****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 11:18:27 1999; MasPar time 8.07 Seconds
60.852 Million cell updates/sec
[redacted]ular output not generated.

Title: >US-08-599-226-3
Description: (1-9) from US08599226.pep
Perfect Score: 66
Sequence: 1 QRYNRAPYX 9

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 22.838; Variance 29.153; scale 0.783

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

[redacted]ult No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	57	86.4	560	5	O44626	K11D12.1 PROTEIN.	3.24e-01
2	54	81.8	248	5	Q27562	PRTC.	1.40e+00
3	54	81.8	260	13	O42265	20S PROTEASOME SUBUNIT	1.40e+00
4	54	81.8	448	5	O02441	GDP-DISSOCIATION INHIB	1.40e+00
5	54	81.8	1277	5	O17517	ZC132.5 PROTEIN.	1.40e+00
6	53	80.3	790	2	Q55956	ABC TRANSPORTER.	2.27e+00
7	52	78.8	90	5	O62346	R11.2 PROTEIN.	3.64e+00
8	52	78.8	494	5	P92024	MEX-1.	3.64e+00
9	52	78.8	509	5	Q21017	COSMID F58H12.	3.64e+00
10	51	77.3	170	2	Q48279	SIMILARITY WITH HEAD C	5.81e+00
11	50	75.8	178	4	Q92592	CD89_L10.	9.23e+00
12	50	75.8	201	4	Q99935	BASIC PROLINE-RICH PRO	9.23e+00
13	50	75.8	332	10	O82438	ANTIFREEZE PROTEIN.	9.23e+00
14	50	75.8	450	2	O50511	ZINC PROTEASE.	9.23e+00
15	50	75.8	521	2	Q44677	NEUTRAL PROTEASE.	9.23e+00
16	50	75.8	3082	14	O36979	POLYPROTEIN.	9.23e+00
17	49	74.2	641	5	Q18586	COSMID C43H6.	1.46e+01
18	48	72.7	43	10	O49246	REVERSE TRANSCRIPTASE	2.28e+01
19	48	72.7	44	10	O49247	REVERSE TRANSCRIPTASE	2.28e+01
20	48	72.7	192	5	Q19008	COSMID D2096.	2.28e+01
21	48	72.7	231	3	O74961	HYPOTHETICAL 25.6 KD P	2.28e+01
22	48	72.7	307	2	O50244	TRANSPOSASE HOMOLOG.	2.28e+01
23	48	72.7	337	2	O87136	ORF139-17 PROTEIN.	2.28e+01
24	48	72.7	337	2	O34232	ORF40X1 PROTEIN.	2.28e+01
25	48	72.7	367	13	Q91905	FORK HEAD PROTEIN.	2.28e+01
26	48	72.7	370	13	Q91904	FORK HEAD PROTEIN.	2.28e+01
27	48	72.7	375	2	O82974	ATP-DEPENDENT PROTEINA	2.28e+01
28	48	72.7	534	13	P70056	FORKHEAD ACTIVIN SIGNA	2.28e+01
29	48	72.7	697	3	O74504	HYPOTHETICAL 77.8 KD P	2.28e+01
30	48	72.7	1032	2	P72637	HYPOTHETICAL 117.2 KD	2.28e+01
31	48	72.7	1495	11	P70587	DENSIN-180.	2.28e+01
32	48	72.7	1496	4	Q92626	MYELOBLAST KIAA0230 (F	2.28e+01
33	48	72.7	2896	5	O61363	HEMOCYANIN G-TYPE SUBU	2.28e+01
34	47	71.2	130	5	Q24955	FMRFAMIDE (FRAGMENT).	3.55e+01
35	47	71.2	355	2	P74732	HYPOTHETICAL 41.6 KD P	3.55e+01
36	47	71.2	425	5	O02282	UNC-101 PROTEIN.	3.55e+01
37	47	71.2	436	5	Q10929	HYPOTHETICAL 47.6 KD P	3.55e+01
38	47	71.2	444	5	Q21449	RABGDI PROTEIN.	3.55e+01
39	47	71.2	445	11	Q61598	GUANOSINE DIPHOSPHATE	3.55e+01
40	47	71.2	445	4	O43928	GDP DISSOCIATION INHIB	3.55e+01
41	47	71.2	515	14	O55733	HYPOTHETICAL 55.3 KD P	3.55e+01
42	47	71.2	523	5	O16922	F10D2.11 PROTEIN.	3.55e+01
43	47	71.2	541	11	Q60447	PUTATIVE GAG PROTEIN.	3.55e+01
44	47	71.2	722	10	O23337	HYPOTHETICAL 82.1 KD P	3.55e+01
45	47	71.2	1312	2	O52485	PROLINE DEHYDROGENASE	3.55e+01

ALIGNMENTS

RESULT 1
ID O44626 PRELIMINARY; PRT; 560 AA.
AC O44626;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE K11D12.1 PROTEIN.
GN K11D12.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA HENKHAUS J., WOHLDMANN P., GILLAM B.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AF039047; G2736445; -.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ SEQUENCE 560 AA; 61136 MW; 5906EB7D CRC32;

Query Match 86.4%; Score 57; DB 5; Length 560;

```
  Best Local Similarity  75.0%;  Pred. No. 3.24e-01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     81 QKYKRAPY 88
          |:|:||||
Qy      1 QRYNRAPY 8


RESULT    2
ID   Q27562      PRELIMINARY;      PRT;   248 AA.
AC   Q27562;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PRTC.
GN   PRTC.
OS   DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC   EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SHAULSKY G., ESCALANTE R., LOOMIS W.F.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S25A; ALSO KNOWN AS THE
CC       PROTEASOME A-TYPE FAMILY.
CC   -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
CC       WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH
CC       ARG, PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT
CC       NEUTRAL OR SLIGHTLY BASIC PH.
CC       THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC ACTIVITY.
CC       THE COMPONENT C3 MAY HAVE A POTENTIAL REGULATORY EFFECT ON
CC       ANOTHER COMPONENT(S) OF THE PROTEASOME COMPLEX THROUGH TYROSINE
CC       PHOSPHORYLATION.
CC   -!- PATHWAY: IS INVOLVED IN AN ATP/UBIQUITIN-DEPENDENT NON-LYSOSOMAL
CC       PROTEOLYTIC PATHWAY.
CC   -!- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
CC       SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: PROTEASOMES ARE FOUND IN THE CYTOPLASM AND
CC       ALSO IN THE NUCLEUS.
DR   EMBL; U60168; G1405364; -.
DR   PROSITE; PS00388; PROTEASOME_A; 1.
DR   PFAM; PF00227; proteasome; 1.
KW   PROTEASOME; HYDROLASE; PROTEASE.
SQ   SEQUENCE   248 AA;  27988 MW;  BEE84ACC CRC32;

  Query Match           81.8%;  Score 54;  DB 5;  Length 248;
  Best Local Similarity  75.0%;  Pred. No. 1.40e+00;
  Matches     6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    121 QRYGRRPY 128
          ||| | ||
Qy      1 QRYNRAPY 8


RESULT    3
ID   O42265      PRELIMINARY;      PRT;   260 AA.
AC   O42265;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   20S PROTEASOME SUBUNIT C2.
GN   CC2.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SINGH I., WAGNER B.J.;
RL   SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF027978; G3136063; -.
DR   PFAM; PF00227; proteasome; 1.
KW   PROTEASOME.
SQ   SEQUENCE   260 AA;  28925 MW;  EF38499F CRC32;

  Query Match           81.8%;  Score 54;  DB 13;  Length 260;
  Best Local Similarity  75.0%;  Pred. No. 1.40e+00;
  Matches     6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    120 QRYGRRPY 127
          ||| | ||
Qy      1 QRYNRAPY 8


RESULT    4
ID   O02441      PRELIMINARY;      PRT;   448 AA.
AC   O02441;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   GDP-DISSOCIATION INHIBITOR.
GN   GDI.
OS   GEODIA CYDONIUM (SPONGE).
OC   EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; TETRACTINOMORPHA;
OC   ASTROPHORIDA; GEODIIDAE; GEODIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KRASKO A., SCHEFFER U., KOZIOL C., PANCER Z., BATEL R., BADRIA F.A.,
RA   MUELLER W.E.G.;
RL   AQUATIC TOXICOL. 37:157-168(1997).
DR   EMBL; X94983; E218570; -.
DR   PFAM; PF00996; GDI; 1.
SQ   SEQUENCE   448 AA;  50215 MW;  848511EC CRC32;

  Query Match           81.8%;  Score 54;  DB 5;  Length 448;
  Best Local Similarity  62.5%;  Pred. No. 1.40e+00;
  Matches     5;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    217 QRYGKSPY 224
          ||| ::||
Qy      1 QRYNRAPY 8


RESULT    5
ID   O17517      PRELIMINARY;      PRT;  1277 AA.
AC   O17517;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ZC132.5 PROTEIN.
GN   ZC132.5.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   BRADSHAW H., DEVLIN K.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF014939; G2275629; -.
SQ   SEQUENCE   1277 AA;  146496 MW;  8A6DB899 CRC32;

  Query Match           81.8%;  Score 54;  DB 5;  Length 1277;
  Best Local Similarity 100.0%;  Pred. No. 1.40e+00;
  Matches    7;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    877 QRYNRAP 883
          |||||||
Qy      1 QRYNRAP 7


RESULT    6
ID   Q55956      PRELIMINARY;      PRT;   790 AA.
AC   Q55956;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
     ABC TRANSPORTER.
     SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC   BACTERIA; CYANOBACTERIA; CHROOCOCCALES; SYNECHOCYSTIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 96127529.
RA   KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA   SUGIURA M., TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. sequence features in the 1Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA RES. 2:153-166(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. PCC6803. II. Sequence determination of the entire
RT   genome and assignment of potential protein-coding regions.";
RL   DNA RES. 3:109-136(1996).
DR   EMBL; D64005; D1011375; -.
DR   PFAM; PF00005; ABC_tran; 1.
DR   PFAM; PF00498; FHA; 2.
SQ   SEQUENCE   790 AA;  87656 MW;  436D8E61 CRC32;

  Query Match           80.3%;  Score 53;  DB 2;  Length 790;
  Best Local Similarity 62.5%;  Pred. No. 2.27e+00;
  Matches    5;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    475 QRYSRSSY 482
          |||:|::|
Qy      1 QRYNRAPY 8


RESULT    7
ID   O62346      PRELIMINARY;      PRT;    90 AA.
AC   O62346;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   R11.2 PROTEIN.
GN   R11.2.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MCMURRAY A.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z81577; E1348694; -.
SQ   SEQUENCE   90 AA;  10937 MW;  80B83F8F CRC32;

  Query Match           78.8%;  Score 52;  DB 5;  Length 90;
  Best Local Similarity 75.0%;  Pred. No. 3.64e+00;
  Matches    6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db     21 QRYDRAIY 28
          |||:|| |
Qy      1 QRYNRAPY 8


RESULT    8
ID   P92024      PRELIMINARY;      PRT;   494 AA.
AC   P92024;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEX-1.
GN   MEX-1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RX   MEDLINE; 97195715.
RA   GUEDES S., PRIESS J.R.;
RT   "The C. elegans MEX-1 protein is present in germline blastomeres and
RT   is a P granule component.";
RL   DEVELOPMENT 124:731-739(1997).
DR   EMBL; U81043; G1899062; -.
DR   PFAM; PF00642; zf-CCCH; 2.
SQ   SEQUENCE   494 AA;  56608 MW;  716506D2 CRC32;

  Query Match           78.8%;  Score 52;  DB 5;  Length 494;
  Best Local Similarity 62.5%;  Pred. No. 3.64e+00;
  Matches    5;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    263 QRYRRPPF 270
          ||| |:|:
Qy      1 QRYNRAPY 8


RESULT    9
ID   Q21017      PRELIMINARY;      PRT;   509 AA.
AC   Q21017;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   COSMID F58H12.
GN   F58H12.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
```

```
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MILLER N.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U40416; G1065490; -.
SQ   SEQUENCE   509 AA;  55953 MW;  D0EC1387 CRC32;

  Query Match           78.8%;  Score 52;  DB 5;  Length 509;
  Best Local Similarity  71.4%;  Pred. No. 3.64e+00;
  Matches     5;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    268 RYQRTPY 274
          || |:||
Qy      2 RYNRAPY 8


RESULT   10
ID   Q48279      PRELIMINARY;      PRT;   170 AA.
AC   Q48279;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   SIMILARITY WITH HEAD COMPLETION/STABILIZATION PROTEIN.
OS   HAEMOPHILUS SOMNUS.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC   HAEMOPHILUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HS25;
RA   PONTAROLLO R.A.;
RL   THESIS (1995), V. I. D. O., UNIVERSITY OF SASKATCHEWAN.
DR   EMBL; U28154; G915368; -.
SQ   SEQUENCE   170 AA;  19923 MW;  91428D5E CRC32;

  Query Match           77.3%;  Score 51;  DB 2;  Length 170;
  Best Local Similarity  75.0%;  Pred. No. 5.81e+00;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    107 QRYKRAVY 114
          |||:|| |
Qy      1 QRYNRAPY 8


RESULT   11
ID   Q92592      PRELIMINARY;      PRT;   178 AA.
AC   Q92592;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   CD89_L10.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   TOYABE S., KUWANO Y., TAKEDA K., UCHIYAMA M., ABO T.;
RT   "Alternatively spliced forms of monocyte IgA Fc receptors in patients
RT   with IgA nephropathy.";
RL   SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D87861; D1014171; -.
SQ   SEQUENCE   178 AA;  19593 MW;  EC5B1F1C CRC32;

  Query Match           75.8%;  Score 50;  DB 4;  Length 178;
  Best Local Similarity  71.4%;  Pred. No. 9.23e+00;
  Matches     5;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     87 WYNRSPY 93
          :|||:||
Qy      2 RYNRAPY 8


RESULT   12
ID   Q99935      PRELIMINARY;      PRT;   201 AA.
AC   Q99935;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   BASIC PROLINE-RICH PROTEIN.
GN   BPLP.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96309097.
RA   DICKINSON D.P., THIESSE M.;
RT   "cDNA cloning of an abundant human lacrimal gland mRNA encoding a
RT   novel tear protein.";
RL   CURR. EYE RES. 15:377-386(1996).
DR   EMBL; S83198; G1836022; -.
SQ   SEQUENCE   201 AA;  22870 MW;  650A8F70 CRC32;

  Query Match           75.8%;  Score 50;  DB 4;  Length 201;
  Best Local Similarity  62.5%;  Pred. No. 9.23e+00;
  Matches     5;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db     22 QRFSRRPY 29
          ||::| ||
Qy      1 QRYNRAPY 8


RESULT   13
ID   O82438      PRELIMINARY;      PRT;   332 AA.
AC   O82438;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ANTIFREEZE PROTEIN.
GN   AFP.
OS   DAUCUS CAROTA (CARROT).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC   ASTERIDAE; ARALIALES; APIACEAE; DAUCUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. AUTUMN KING; TISSUE=TAP ROOT;
RX   MEDLINE; 98429644.
RA   WORRALL D., ELIAS L., ASHFORD D., SMALLWOOD M., SIDEBOTTOM C.,
RA   LILLFORD P., TELFORD J., HOLT C., BOWLES D.;
RT   "A carrot leucine-rich-repeat protein that inhibits ice
RT   recrystallization.";
RL   SCIENCE 282:115-117(1998).
DR   EMBL; AF055480; G3702803; -.
SQ   SEQUENCE   332 AA;  36845 MW;  67194791 CRC32;
```

```
  Query Match           75.8%;  Score 50;  DB 10;  Length 332;
  Best Local Similarity  50.0%;  Pred. No. 9.23e+00;
  Matches     4;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    311 QRFDRTAY 318
          ||::|::|
Qy      1 QRYNRAPY 8


RESULT   14
ID   O50511      PRELIMINARY;      PRT;   450 AA.
AC   O50511;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ZINC PROTEASE.
GN   SC9B10.04.
OS   STREPTOMYCES COELICOLOR.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; STREPTOMYCINEAE; STREPTOMYCETACEAE; STREPTOMYCES.
     [1]
     SEQUENCE FROM N.A.
     STRAIN=A3(2);
RA   OLIVER K., HARRIS D.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE; 97000351.
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   MOL. MICROBIOL. 21:77-96(1996).
DR   EMBL; AL009204; E1202335; -.
KW   PROTEASE.
SQ   SEQUENCE   450 AA;  49045 MW;  3E30B71A CRC32;

  Query Match           75.8%;  Score 50;  DB 2;  Length 450;
  Best Local Similarity  62.5%;  Pred. No. 9.23e+00;
  Matches     5;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    147 QRYDNVPY 154
          |||:  ||
Qy      1 QRYNRAPY 8


RESULT   15
ID   Q44677      PRELIMINARY;      PRT;   521 AA.
AC   Q44677;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NEUTRAL PROTEASE.
OS   BACILLUS AMYLOLIQUEFACIENS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SHIMADA H., HONJO M., MITA I., NAKAYAMA A., AKAOKA A., MANABE K.,
RA   FURUTANI Y.;
RL   J. BIOTECHNOL. 2:75-85(1985).
DR   EMBL; M36723; G143353; -.
DR   PFAM; PF00099; zn-protease; 1.
KW   PROTEASE.
SQ   SEQUENCE   521 AA;  56725 MW;  64AFFE5F CRC32;

  Query Match           75.8%;  Score 50;  DB 2;  Length 521;
  Best Local Similarity  62.5%;  Pred. No. 9.23e+00;
  Matches     5;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    308 QKYNRNSY 315
          |:||| :|
Qy      1 QRYNRAPY 8


Search completed: Thu Sep  2 11:18:44 1999
Job time : 17 secs.
```

****

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 11:21:51 1999; MasPar time 4.46 Seconds
146.856 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-4
Description: (1-12) from US08599226.pep
Perfect Score: 66
Sequence: 1 VSYLSTASSLDX 12

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 23.196; Variance 28.631; scale 0.810

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	55	83.3	649	3	O60167	PROTEIN COMPLEX ASSEMB	8.09e-01
2	51	77.3	236	3	Q12282	CHROMOSOME XV READING	5.65e+00
3	51	77.3	3623	11	O70244	INTRINSIC FACTOR-B12 R	5.65e+00
4	50	75.8	293	2	O51095	CONSERVED HYPOTHETICAL	9.06e+00
5	50	75.8	705	2	P94888	CADMIUM RESISTANCE PRO	9.06e+00
6	49	74.2	599	2	O84716	PHOSPHOENOLPYRUVATE CA	1.44e+01
7	48	72.7	292	14	O55748	PUTATIVE RIBONUCLEASE.	2.28e+01
8	48	72.7	467	4	O43298	KIAA0414.	2.28e+01
9	48	72.7	832	4	Q92831	P/CAF.	2.28e+01
10	48	72.7	1245	5	O16568	F40E12.2 PROTEIN (FRAG	2.28e+01
11	47	71.2	127	14	O11696	NUCLEOPROTEIN (FRAGMEN	3.58e+01
12	47	71.2	149	11	Q61836	MYELIN BASIC PROTEIN (	3.58e+01
13	47	71.2	195	11	Q01585	MYELIN BASIC PROTEIN (	3.58e+01
14	47	71.2	197	4	Q15339	GOLLI-MBP.	3.58e+01
15	47	71.2	239	2	Q00510	SRMX PROTEIN.	3.58e+01
16	47	71.2	250	11	Q03139	MYELIN BASIC PROTEIN (	3.58e+01
17	47	71.2	255	10	O23112	AP2 DOMAIN CONTAINING	3.58e+01
18	47	71.2	303	14	Q84594	SIMILAR TO PBCV-1 SERI	3.58e+01
19	47	71.2	304	4	Q15340	GOLLI-MBP.	3.58e+01
20	47	71.2	458	4	O60326	KIAA0582 PROTEIN (FRAG	3.58e+01
21	47	71.2	959	5	O18359	PUTATIVE KRUEPPEL TARG	3.58e+01
22	47	71.2	1128	1	O52009	REPJ PROTEIN.	3.58e+01
23	47	71.2	1128	1	O51999	REPI PROTEIN.	3.58e+01
24	46	69.7	133	2	O84670	HYPOTHETICAL 14.7 KD P	5.58e+01
25	46	69.7	258	5	Q19850	COSMID F28B12.	5.58e+01
26	46	69.7	296	14	Q69118	HYPOTHETICAL PROTEIN (	5.58e+01
27	46	69.7	500	2	O25869	ANTHRANILATE SYNTHASE	5.58e+01
28	46	69.7	503	5	Q26594	AMINO ACID PERMEASE.	5.58e+01
29	46	69.7	1022	6	Q28628	A-KINASE ANCHORING PRO	5.58e+01
30	46	69.7	2910	5	Q26008	RNA POLYMERASE I.	5.58e+01
31	45	68.2	121	14	O90626	GLYCOPROTEIN GJ.	8.64e+01
32	45	68.2	131	14	O90638	GLYCOPROTEIN GJ.	8.64e+01
33	45	68.2	164	5	Q23378	COSMID ZC64.	8.64e+01
34	45	68.2	261	1	O59473	261AA LONG HYPOTHETICA	8.64e+01
35	45	68.2	276	8	Q33658	ATP SYNTHASE A CHAIN (	8.64e+01
36	45	68.2	370	10	O65492	HYPOTHETICAL 42.1 KD P	8.64e+01
37	45	68.2	384	2	Q11197	HYPOTHETICAL 41.9 KD P	8.64e+01
38	45	68.2	448	5	O77045	NWEG PRECURSOR (EC 3.2	8.64e+01
39	45	68.2	471	10	O23552	HYPOTHETICAL 52.8 KD P	8.64e+01
40	45	68.2	546	2	O69236	MERCURIC REDUCTASE.	8.64e+01
41	45	68.2	608	10	P92973	CCA1.	8.64e+01
42	45	68.2	619	4	Q13042	CDC16HS.	8.64e+01
43	45	68.2	631	2	O86216	MERCURIC REDUCTASE.	8.64e+01
44	45	68.2	650	2	O50417	MULTI-FUNCTIONAL ENZYM	8.64e+01
45	45	68.2	1369	13	P79950	TYROSINE KINASE RECEPT	8.64e+01

ALIGNMENTS

RESULT 1
ID O60167 PRELIMINARY; PRT; 649 AA.
AC O60167;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PROTEIN COMPLEX ASSEMBLY PROTEIN.
GN SPBC19F8.03C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA BECK A., REINHARDT R., WOOD V., RAJANDREAM M.A., BARRELL B.G.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AL023594; E1293401; -.
SQ SEQUENCE 649 AA; 72985 MW; 9C207DB2 CRC32;

Query Match 83.3%; Score 55; DB 3; Length 649;
Best Local Similarity 63.6%; Pred. No. 8.09e-01;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Db    233 INYLSTARSLE 243
          ::|||||  ||:
Qy      1 VSYLSTASSLD 11
```

RESULT 2
ID Q12282 PRELIMINARY; PRT; 236 AA.
AC Q12282;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE CHROMOSOME XV READING FRAME ORF YOR214C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA BOYER J., FAIRHEAD C., GAILLON L., GALISSON F., MICHAUX G.,
RA THIERRY A., DUJON B.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.

```
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FY1679;
RA   GALISSON F., DUJON B.;
RL   SUBMITTED (OCT-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z75122; E252081; -.
DR   EMBL; X92441; G1050766; -.
SQ   SEQUENCE   236 AA;  26156 MW;  9759C243 CRC32;

  Query Match           77.3%;  Score 51;  DB 3;  Length 236;
  Best Local Similarity  60.0%;  Pred. No. 5.65e+00;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    175 ISYLSSTTSL 184
          :||||:::||
Qy      1 VSYLSTASSL 10
```

```
RESULT    3
ID   O70244      PRELIMINARY;      PRT;  3623 AA.
AC   O70244;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   INTRINSIC FACTOR-B12 RECEPTOR PRECURSOR.
GN   CUBILIN.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98148073.
RA   MOESTRUP S.K., KOZYRAKI R., KRISTIANSEN M., KAYSEN J.H.,
RA   RASMUSSEN H.H., BRAULT D., PONTILLON F., GODA F.O., CHRISTENSEN E.I.,
RA   HAMMOND T.G., VERROUST P.J.;
RT   "The intrinsic factor-vitamin B12 receptor and target of teratogenic
RT   antibodies is a megalin-binding peripheral membrane protein with
RT   homology to developmental proteins.";
RL   J. BIOL. CHEM. 273:5235-5242(1998).
DR   EMBL; AF022247; G3834380; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS01187; EGF_CA; 4.
KW   SIGNAL; GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   3623       INTRINSIC FACTOR-B12 RECEPTOR.
SQ   SEQUENCE   3623 AA;  398981 MW;  ADF804DC CRC32;

  Query Match           77.3%;  Score 51;  DB 11;  Length 3623;
  Best Local Similarity  54.5%;  Pred. No. 5.65e+00;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    697 ITYLTTQSDLD 707
          ::||:| | ||
Qy      1 VSYLSTASSLD 11
```

```
RESULT    4
ID   O51095      PRELIMINARY;      PRT;   293 AA.
AC   O51095;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   BB0068.
OS   BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC   BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE; 98065943.
RA   FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA   LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA   DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA   PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA   VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA   UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA   GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA   SMITH H.O., VENTER J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   NATURE 390:580-586(1997).
DR   EMBL; AE001120; G2687956; -.
DR   TIGR; BB0068; -.
SQ   SEQUENCE   293 AA;  33278 MW;  3FBBB9E2 CRC32;

  Query Match           75.8%;  Score 50;  DB 2;  Length 293;
  Best Local Similarity  60.0%;  Pred. No. 9.06e+00;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    198 AYLSTPNSLE 207
          :||||::||:
Qy      2 SYLSTASSLD 11
```

```
RESULT    5
ID   P94888      PRELIMINARY;      PRT;   705 AA.
AC   P94888;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CADMIUM RESISTANCE PROTEIN.
GN   CADA.
OS   LACTOCOCCUS LACTIS (STREPTOCOCCUS LACTIS).
OG   PLASMID PND302.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC   LACTOCOCCUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M71;
RA   LIU C.Q., CHIA G.L., DUNN N.W.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U78967; G1699049; -.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
DR   PFAM; PF00122; E1-E2_ATPase; 1.
DR   PFAM; PF00403; HMA; 1.
KW   PLASMID; HYDROLASE; TRANSMEMBRANE; PHOSPHORYLATION; ATP-BINDING.
FT   MOD_RES     398    398       PHOSPHORYLATION (PROBABLE).
SQ   SEQUENCE   705 AA;  76466 MW;  6648C08D CRC32;

  Query Match           75.8%;  Score 50;  DB 2;  Length 705;
  Best Local Similarity  54.5%;  Pred. No. 9.06e+00;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    422 INYLSIISSLE 432
          ::|||  |||:
Qy      1 VSYLSTASSLD 11
```

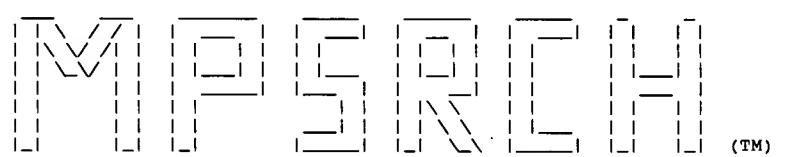
```
RESULT    6
ID   O84716      PRELIMINARY;      PRT;   599 AA.
AC   O84716;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PHOSPHOENOLPYRUVATE CARBOXYKINASE.
GN   PCKA.
OS   CHLAMYDIA TRACHOMATIS.
OC   BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
```

```
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RT   "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT   Chlamydia trachomatis.";
RL   SCIENCE 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RL   SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AE001341; G3329165; -.
DR   PROSITE; PS00505; PEPCK_GTP; 1.
KW   PYRUVATE.
SQ   SEQUENCE   599 AA;  66244 MW;  8DE1FE30 CRC32;

  Query Match           74.2%;  Score 49;  DB 2;  Length 599;
  Best Local Similarity  54.5%;  Pred. No. 1.44e+01;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    533 VGYLPTAEGLN 543
          |:||:|| :|:
Qy      1 VSYLSTASSLD 11


RESULT    7
ID   O55748      PRELIMINARY;      PRT;   292 AA.
AC   O55748;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE   PUTATIVE RIBONUCLEASE.
OS   CHILO IRIDESCENT VIRUS (CIV) (INSECT IRIDESCENT VIRUS TYPE 6).
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; IRIDOVIRIDAE; IRIDOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BAHR U., TIDONA C.A., DARAI G.;
RL   VIRUS GENES 0:0-0(1997).
DR   EMBL; AF003534; G2738432; -.
SQ   SEQUENCE   292 AA;  33664 MW;  706D89A2 CRC32;

  Query Match           72.7%;  Score 48;  DB 14;  Length 292;
  Best Local Similarity  45.5%;  Pred. No. 2.28e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     39 IPYLTTPESLN 49
          ::||:|: ||:
Qy      1 VSYLSTASSLD 11


RESULT    8
ID   O43298      PRELIMINARY;      PRT;   467 AA.
AC   O43298;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   KIAA0414.
GN   KIAA0414 OR ZNF-X.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   ISHIKAWA K., NAGASE T., NAKAJIMA D., SEKI N., OHIRA M., MIYAJIMA N.,
RA   TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   ZHANG X., LIU C.-X., LISITSINA M.N., MUSCO S., LISITSYN N.A.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AB007874; D1025766; -.
DR   EMBL; AF049907; G2952308; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE   467 AA;  52630 MW;  FA610780 CRC32;

  Query Match           72.7%;  Score 48;  DB 4;  Length 467;
  Best Local Similarity  63.6%;  Pred. No. 2.28e+01;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    101 VSYLTAASFLQ 111
          ||||::|| |:
Qy      1 VSYLSTASSLD 11


RESULT    9
ID   Q92831      PRELIMINARY;      PRT;   832 AA.
AC   Q92831;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   P/CAF.
GN   P/CAF.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 96300317.
RA   YANG X.J., OGRYZKO V.V., NISHIKAWA J., HOWARD B.H., NAKATANI Y.;
RT   "A p300/CBP-associated factor that competes with the adenoviral
RT   oncoprotein E1A.";
RL   NATURE 382:319-324(1996).
DR   EMBL; U57317; G1491937; -.
DR   PFAM; PF00439; bromodomain; 1.
DR   PFAM; PF00583; Acetyltransf; 1.
SQ   SEQUENCE   832 AA;  92926 MW;  C891243F CRC32;

  Query Match           72.7%;  Score 48;  DB 4;  Length 832;
  Best Local Similarity  63.6%;  Pred. No. 2.28e+01;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    394 ISYNSTSSSLE 404
          :|| ||:|||:
Qy      1 VSYLSTASSLD 11


RESULT   10
ID   O16568      PRELIMINARY;      PRT;  1245 AA.
AC   O16568;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F40E12.2 PROTEIN (FRAGMENT).
GN   F40E12.2.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
```

```
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   GOELA D.;
RL   SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF016659; G2315574; -.
DR   PFAM; PF00023; ank; 1.
DR   PFAM; PF00533; BRCT; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1245 AA;  142852 MW;  A3C77849 CRC32;

  Query Match           72.7%;  Score 48;  DB 5;  Length 1245;
  Best Local Similarity  45.5%;  Pred. No. 2.28e+01;
  Matches    5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    290 VDFLTSANSLE 300
          | :|::|:||:
Qy      1 VSYLSTASSLD 11


RESULT   11
ID   O11696      PRELIMINARY;      PRT;   127 AA.
AC   O11696;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NUCLEOPROTEIN (FRAGMENT).
OS   MEASLES VIRUS (SUBACUTE SCLEROSE PANENCEPHALITIS VIRUS).
OC   VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC   PARAMYXOVIRIDAE; PARAMYXOVIRINAE; MORBILLIVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=92-E;
RX   MEDLINE; 97278133.
RA   YAMAGUCHI S.;
RT   "Identification of three lineages of wild measles virus by nucleotide
RT   sequence analysis of N, P, M, F, and L genes in Japan.";
RL   J. MED. VIROL. 52:113-120(1997).
DR   EMBL; D87487; D1020995; -.
KW   NUCLEOPROTEIN.
FT   NON_TER       1      1
SQ   SEQUENCE   127 AA;  13950 MW;  42D75A2C CRC32;

  Query Match           71.2%;  Score 47;  DB 14;  Length 127;
  Best Local Similarity  50.0%;  Pred. No. 3.58e+01;
  Matches    5;  Conservative    5;  Mismatches   0;  Indels   0;  Gaps   0;

Db     67 AYLPTSTPLD 76
          :||:|:::||
Qy      2 SYLSTASSLD 11


RESULT   12
ID   Q61836      PRELIMINARY;      PRT;   149 AA.
AC   Q61836;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MYELIN BASIC PROTEIN (FRAGMENT).
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=BRAIN;
RX   MEDLINE; 87118269.
RA   NEWMAN S., KITAMURA K., CAMPAGNONI A.T.;
RT   "Identification of a cDNA coding for a fifth form of myelin basic
RT   protein in mouse.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 84:886-890(1987).
DR   EMBL; M15062; G199051; -.
DR   PROSITE; PS00569; MYELIN_MBP; 1.
KW   ALTERNATIVE SPLICING; MYELIN.
FT   NON_TER       1      1
SQ   SEQUENCE   149 AA;  16226 MW;  A6DC1599 CRC32;

  Query Match           71.2%;  Score 47;  DB 11;  Length 149;
  Best Local Similarity  66.7%;  Pred. No. 3.58e+01;
  Matches    6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db      4 YLATASTMD 12
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT   13
ID   Q01585      PRELIMINARY;      PRT;   195 AA.
AC   Q01585;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MYELIN BASIC PROTEIN (MBP).
GN   MBP.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57 BLACK; TISSUE=BONE MARROW;
RX   MEDLINE; 93057537.
RA   GRIMA B., ZELENIKA D., PESSAC B.;
RT   "A novel transcript overlapping the myelin basic protein gene.";
RL   J. NEUROCHEM. 59:2318-2323(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=BRAIN;
RX   MEDLINE; 93186801.
RA   CAMPAGNONI A.T., PRIBYL T.M., CAMPAGNONI C.W., KAMPF K.,
RA   AMUR-UMARJEE S., LANDRY C., HANDLEY V., NEWMAN S., GARBAY B.,
RA   KITAMURA K.;
RT   "Structure and developmental regulation of Golli-mbp, a 105-kilobase
RT   gene that encompasses the myelin basic protein gene and is expressed
RT   in cells in the oligodendrocyte lineage in the brain.";
RL   J. BIOL. CHEM. 268:4930-4938(1993).
CC   -!- FUNCTION: THIS PROTEIN MAY FUNCTION TO MAINTAIN PROPER STRUCTURE
CC       OF MYELIN.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC SIDE OF MYELIN.
DR   EMBL; X67319; G51333; -.
DR   EMBL; L07508; G193587; -.
DR   MGD; MGI:96925; MBP.
KW   MYELIN; STRUCTURAL PROTEIN; ACETYLATION; METHYLATION; PHOSPHORYLATION;
KW   AUTOIMMUNE ENCEPHALOMYELITIS; ALTERNATIVE SPLICING.
SQ   SEQUENCE   195 AA;  21004 MW;  557D83EA CRC32;

  Query Match           71.2%;  Score 47;  DB 11;  Length 195;
  Best Local Similarity  66.7%;  Pred. No. 3.58e+01;
  Matches    6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    146 YLATASTMD 154
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT   14
ID   Q15339      PRELIMINARY;      PRT;   197 AA.
AC   Q15339;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```

```
DE   GOLLI-MBP.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 94068468.
RA   PRIBYL T.M., CAMPAGNONI C.W., KAMPF K., KASHIMA T., HANDLEY V.W.,
RA   MCMAHON J., CAMPAGNONI A.T.;
RT   "The human myelin basic protein gene is included within a
RT   179-kilobase transcription unit: expression in the immune and central
RT   nervous systems.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:10695-10699(1993).
DR   EMBL; L18865; G435060; -.
SQ   SEQUENCE   197 AA;  21522 MW;  D8A8A5BF CRC32;

  Query Match           71.2%;  Score 47;  DB 4;  Length 197;
  Best Local Similarity  66.7%;  Pred. No. 3.58e+01;
  Matches    6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

      148 YLATASTMD 156
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT   15
ID   Q00510      PRELIMINARY;      PRT;   239 AA.
AC   Q00510;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SRMX PROTEIN.
GN   SRMX.
OS   STREPTOMYCES AMBOFACIENS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; STREPTOMYCINEAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BES2281;
RX   MEDLINE; 92374852.
RA   GEISTLICH M., LOSICK R., TURNER J.R., RAU R.N.;
RT   "Characterization of a novel regulatory gene governing the expression
RT   of a polyketide synthase gene in Streptomyces ambofaciens.";
RL   MOL. MICROBIOL. 6:2019-2029(1992).
DR   EMBL; X63451; G46702; -.
SQ   SEQUENCE   239 AA;  26493 MW;  45A8C51A CRC32;

  Query Match           71.2%;  Score 47;  DB 2;  Length 239;
  Best Local Similarity  36.4%;  Pred. No. 3.58e+01;
  Matches    4;  Conservative    5;  Mismatches   2;  Indels   0;  Gaps   0;

Db    109 IGYLETTADLE 119
          ::|| |:: |:
Qy      1 VSYLSTASSLD 11


Search completed: Thu Sep  2 11:22:14 1999
Job time : 23 secs.
```

****



****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

[redacted] on: Thu Sep 2 11:21:25 1999; MasPar time 2.25 Seconds
150.446 Million cell updates/sec
[redacted]ular output not generated.

Title: >US-08-599-226-4
Description: (1-12) from US08599226.pep
Perfect Score: 66
Sequence: 1 VSYLSTASSLDX 12

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 24.203; Variance 24.721; scale 0.979

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	55	83.3	352	1	YDH3_SCHPO	HYPOTHETICAL 39.7 KD P	1.07e-01
2	52	78.8	547	1	MERA_STAAU	MERCURIC REDUCTASE (EC	5.87e-01
3	49	74.2	132	1	YW07_MYCTU	HYPOTHETICAL 14.7 KD P	3.00e+00
4	49	74.2	468	1	YPS7_CAEEL	PUTATIVE SERINE/THREON	3.00e+00
5	49	74.2	471	1	UFO1_MAIZE	FLAVONOL 3-O-GLUCOSYLT	3.00e+00
6	49	74.2	471	1	UFO2_MAIZE	FLAVONOL 3-O-GLUCOSYLT	3.00e+00
7	49	74.2	471	1	UFO3_MAIZE	FLAVONOL 3-O-GLUCOSYLT	3.00e+00
8	49	74.2	2464	1	MAPB_MOUSE	MICROTUBULE-ASSOCIATED	3.00e+00
9	49	74.2	2468	1	MAPB_HUMAN	MICROTUBULE-ASSOCIATED	3.00e+00
10	47	71.2	127	1	MBP_RAT	MYELIN BASIC PROTEIN S	8.55e+00
11	47	71.2	167	1	MBP_CAVPO	MYELIN BASIC PROTEIN (	8.55e+00
12	47	71.2	171	1	MBP_PANTR	MYELIN BASIC PROTEIN (	8.55e+00
13	47	71.2	194	1	MBP_MOUSE	MYELIN BASIC PROTEIN (	8.55e+00
14	47	71.2	196	1	MBP_HUMAN	MYELIN BASIC PROTEIN (	8.55e+00
15	47	71.2	259	1	RNPH_MYCTU	PROBABLE RIBONUCLEASE	8.55e+00
16	47	71.2	259	1	RNPH_MYCLE	PROBABLE RIBONUCLEASE	8.55e+00
17	47	71.2	1107	1	YJEP_ECOLI	HYPOTHETICAL 123.8 KD	8.55e+00
18	46	69.7	152	1	PSAL_GUITH	PHOTOSYSTEM I REACTION	1.42e+01
19	46	69.7	474	1	MERA_STRLI	MERCURIC REDUCTASE (EC	1.42e+01
20	46	69.7	485	1	G6PD_ZYMMO	GLUCOSE-6-PHOSPHATE 1-	1.42e+01
21	46	69.7	541	1	TRPE_VIBPA	ANTHRANILATE SYNTHASE	1.42e+01
22	46	69.7	554	1	NBL4_MOUSE	NBL4 PROTEIN.	1.42e+01
23	46	69.7	577	1	PYRH_YEAST	CTP SYNTHASE 2 (EC 6.3	1.42e+01
24	46	69.7	687	1	VIUA_VIBCH	VIBRIOBACTIN RECEPTOR	1.42e+01
25	46	69.7	753	1	YBHJ_ECOLI	HYPOTHETICAL 81.5 KD P	1.42e+01
26	45	68.2	87	1	YIEC_ERWCH	POTENTIAL OUTER MEMBRA	2.35e+01
27	45	68.2	416	1	YEIJ_ECOLI	HYPOTHETICAL 43.4 KD P	2.35e+01
28	45	68.2	505	1	RADA_SYNY3	DNA REPAIR PROTEIN RAD	2.35e+01
29	45	68.2	565	1	HEMA_IAGUA	HEMAGGLUTININ PRECURSO	2.35e+01
30	45	68.2	566	1	HEMA_IAGU2	HEMAGGLUTININ PRECURSO	2.35e+01
31	45	68.2	631	1	MERA_BACSR	MERCURIC REDUCTASE (EC	2.35e+01
32	45	68.2	633	1	YIJ2_YEAST	HYPOTHETICAL 71.0 KD P	2.35e+01
33	45	68.2	926	1	PTN4_HUMAN	PROTEIN-TYROSINE PHOSP	2.35e+01
34	45	68.2	1545	1	YMH2_YEAST	HYPOTHETICAL 171.1 KD P	2.35e+01
35	44	66.7	169	1	MBP_BOVIN	MYELIN BASIC PROTEIN (	3.83e+01
36	44	66.7	336	1	UL34_EBV	BFRF1 PROTEIN.	3.83e+01
37	44	66.7	341	1	RECA_LACLA	RECA PROTEIN, PLASMID.	3.83e+01
38	44	66.7	421	1	EPC_MOUSE	IG EPSILON CHAIN C REG	3.83e+01
39	44	66.7	429	1	EPC_RAT	IG EPSILON CHAIN C REG	3.83e+01
40	44	66.7	1451	1	A2M2_MOUSE	MURINOGLOBULIN 2 PRECU	3.83e+01
41	43	65.2	256	1	YHI2_YEAST	HYPOTHETICAL 29.0 KD P	6.19e+01
42	43	65.2	312	1	YCE9_YEAST	HYPOTHETICAL 35.6 KD P	6.19e+01
43	43	65.2	333	1	YASD_MYCSM	HYPOTHETICAL 35.9 KD P	6.19e+01
44	43	65.2	350	1	ENP2_TORCA	ELECTROMOTOR NEURON-AS	6.19e+01
45	43	65.2	2209	1	Y166_HUMAN	HYPOTHETICAL PROTEIN K	6.19e+01

ALIGNMENTS

RESULT 1
ID YDH3_SCHPO STANDARD; PRT; 352 AA.
AC Q92348;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 39.7 KD PROTEIN C6G9.03C IN CHROMOSOME I.
GN SPAC6G9.03C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., CONNOR R.E.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- SIMILARITY: SOME, TO YEAST YNL206C.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; Z81317; E276610; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 352 AA; 39679 MW; D92A9357 CRC32;

Query Match 83.3%; Score 55; DB 1; Length 352;
Best Local Similarity 72.7%; Pred. No. 1.07e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Db    302 VSYLNEASSLE 312
          |||| |||| :
Qy      1 VSYLSTASSLD 11
```

RESULT 2
ID MERA_STAAU STANDARD; PRT; 547 AA.
AC P08663;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE MERCURIC REDUCTASE (EC 1.16.1.1) (HG(II) REDUCTASE).

```
GN   MERA.
OS   STAPHYLOCOCCUS AUREUS.
OG   PLASMID PI258.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   STAPHYLOCOCCUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87260937.
RA   LADDAGA R.A., CHU L., MISRA T.K., SILVER S.;
RT   "Nucleotide sequence and expression of the mercurial-resistance
RT   operon from Staphylococcus aureus plasmid pI258.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 84:5106-5110(1987).
CC   -!- FUNCTION: RESISTANCE TO HG(2+) IN BACTERIA APPEARS TO BE GOVERNED
CC       BY A SPECIALIZED SYSTEM WHICH INCLUDES MERCURIC REDUCTASE. MERA
CC       PROTEIN IS RESPONSIBLE FOR VOLATILIZING MERCURY AS HG(0).
CC   -!- CATALYTIC ACTIVITY: HG + NADP(+) + H(+) = HG(2+) + NADPH.
CC   -!- COFACTOR: FAD.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC       OXIDOREDUCTASES CLASS-I.
CC   -!- SIMILARITY: CONTAINS A COPY OF THE HEAVY-METAL-ASSOCIATED (HMA)
CC       DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29436; G459907; -.
DR   PIR; E29504; E29504.
DR   PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
DR   PROSITE; PS01047; HMA; 1.
DR   PFAM; PF00070; pyr_redox; 1.
DR   PFAM; PF00403; HMA; 1.
DR   HSSP; P11959; 1EBD.
KW   MERCURIC RESISTANCE; OXIDOREDUCTASE; FLAVOPROTEIN; FAD; NADP;
KW   MERCURY; REDOX-ACTIVE CENTER; METAL-BINDING; PLASMID.
FT   DOMAIN       10     38       HMA.
FT   NP_BIND      87     117      FAD (ADP PART) (PROBABLE).
FT   DISULFID     123    128      REDOX-ACTIVE.
FT   NP_BIND      378    388      FAD (FLAVIN PART) (BY SIMILARITY).
FT   METAL        544    544      HG(2+) (POTENTIAL).
FT   METAL        545    545      HG(2+) (POTENTIAL).
SQ   SEQUENCE   547 AA;  58565 MW;  60DAFC29 CRC32;

  Query Match           78.8%;  Score 52;  DB 1;  Length 547;
  Best Local Similarity  45.5%;  Pred. No. 5.87e-01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    235 VDYLTSTSALE 245
          | ||:::|:|:
Qy      1 VSYLSTASSLD 11


RESULT    3
ID   YW07_MYCTU     STANDARD;      PRT;   132 AA.
AC   Q10847;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 14.7 KD PROTEIN CY39.07C.
GN   MTCY39.07C.
OS   MYCOBACTERIUM TUBERCULOSIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: TO M.TUBERCULOSIS MTCY48.04C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74025; E248788; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   132 AA;  14731 MW;  3D577F11 CRC32;

  Query Match           74.2%;  Score 49;  DB 1;  Length 132;
  Best Local Similarity  60.0%;  Pred. No. 3.00e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db      9 IAYLSTSESL 18
          ::||||: ||
Qy      1 VSYLSTASSL 10


RESULT    4
ID   YPS7_CAEEL     STANDARD;      PRT;   468 AA.
AC   Q20085;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PUTATIVE SERINE/THREONINE-PROTEIN KINASE F35H8.7 IN CHROMOSOME II
DE   (EC 2.7.1.-).
GN   F35H8.7.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   BERKS M.;
RL   SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   REVISIONS.
RC   STRAIN=BRISTOL N2;
RA   JONES S.J.M.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC       PROTEIN KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z36752; E1346252; -.
DR   WORMPEP; F35H8.7; CE09940.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 2.
DR   HSSP; P08631; 2HCK.
KW   HYPOTHETICAL PROTEIN; TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE;
KW   ATP-BINDING.
FT   DOMAIN       106    357      PROTEIN KINASE.
FT   NP_BIND      112    121      ATP (BY SIMILARITY).
FT   BINDING      136    136      ATP (BY SIMILARITY).
FT   ACT_SITE     224    224      BY SIMILARITY.
SQ   SEQUENCE   468 AA;  54045 MW;  480A5814 CRC32;

  Query Match           74.2%;  Score 49;  DB 1;  Length 468;
```

```
  Best Local Similarity  60.0%;  Pred. No. 3.00e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    379 ISYLTTESPL 388
          :|||:| |:|
Qy      1 VSYLSTASSL 10


RESULT    5
ID   UFO1_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16166;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MCC ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88284304.
RA   RALSTON E.J., ENGLISH J.J., DOONER H.K.;
RT   "Sequence of three bronze alleles of maize and correlation with the
RT   genetic fine structure.";
RL   GENETICS 119:185-197(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13500; G1030071; -.
DR   EMBL; X07940; G22205; -.
DR   PIR; S01052; S01052.
DR   PIR; S08324; S08324.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48769 MW;  8AE03FD2 CRC32;

  Query Match           74.2%;  Score 49;  DB 1;  Length 471;
  Best Local Similarity  80.0%;  Pred. No. 3.00e+00;
  Matches     8;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     46 LSFLSTASSL 55
          :|:|||||||
Qy      1 VSYLSTASSL 10


RESULT    6
ID   UFO2_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16165;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MC2 ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13501; G295854; -.
DR   PIR; S08325; S08325.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48621 MW;  3158C5E0 CRC32;

  Query Match           74.2%;  Score 49;  DB 1;  Length 471;
  Best Local Similarity  80.0%;  Pred. No. 3.00e+00;
  Matches     8;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     46 LSFLSTASSL 55
          :|:|||||||
Qy      1 VSYLSTASSL 10


RESULT    7
ID   UFO3_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16167;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-W22 ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88284304.
RA   RALSTON E.J., ENGLISH J.J., DOONER H.K.;
RT   "Sequence of three bronze alleles of maize and correlation with the
RT   genetic fine structure.";
RL   GENETICS 119:185-197(1988).
```

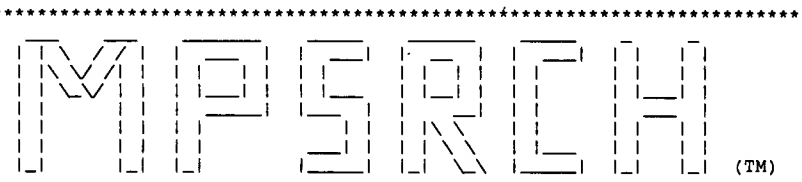
```
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13502; G22506; -.
DR   EMBL; X07937; G22210; -.
DR   PIR; S01037; S01037.
DR   PIR; S08326; S08326.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48673 MW;  4A3C6193 CRC32;

  Query Match           74.2%;  Score 49;  DB 1;  Length 471;
  Best Local Similarity  80.0%;  Pred. No. 3.00e+00;
  Matches    8;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     46 LSFLSTASSL 55
          :|:|||||||
Qy      1 VSYLSTASSL 10


RESULT    8
ID   MAPB_MOUSE     STANDARD;      PRT;  2464 AA.
AC   P14873;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   MICROTUBULE-ASSOCIATED PROTEIN 1B (MAP1.2) (MAP1(X)) [CONTAINS: LIGHT
DE   CHAIN LC1].
GN   MAP1B OR MTAP5.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER; TISSUE=BRAIN;
RX   MEDLINE; 90094539.
RA   NOBLE M., LEWIS S.A., COWAN N.J.;
RT   "The microtubule binding domain of microtubule-associated protein
RT   MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT   and tau.";
RL   J. CELL BIOL. 109:3367-3376(1989).
CC   -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC       PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC       THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC       TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC       MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC       STABILIZING MICROTUBULES.
CC   -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC       WITH MAP1A AND MAP1B PROTEINS.
CC   -!- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE
CC       KKEE AND KKEI/V, REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER
CC       REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES
CC       BOTH IN VITRO AND IN VIVO.
CC   -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC       OF MAP1B.
CC   -!- SIMILARITY: TO NEURAXIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51396; G53000; -.
DR   PIR; S07549; QRMSP1.
DR   MGD; MGI:97179; MTAP5.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 7.
DR   PFAM; PF00414; MAP1B_neuraxin; 10.
KW   MICROTUBULES; REPEAT; PHOSPHORYLATION.
FT   CHAIN         ?   2464       MAP1B LC1.
FT   DOMAIN      589    787       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                KKEE AND KKEI/V REPEATS).
FT   DOMAIN     1865   2068       12 X 17 AA TANDEM REPEATS.
FT   REPEAT     1865   1881       1.
FT   REPEAT     1882   1898       2.
FT   REPEAT     1899   1915       3.
FT   REPEAT     1916   1932       4.
FT   REPEAT     1933   1949       5.
FT   REPEAT     1950   1966       6.
FT   REPEAT     1967   1983       7.
FT   REPEAT     1984   2000       8.
FT   REPEAT     2001   2017       9.
FT   REPEAT     2018   2034       10.
FT   REPEAT     2035   2051       11.
FT   REPEAT     2052   2068       12.
SQ   SEQUENCE   2464 AA;  270408 MW;  44FFF677 CRC32;

  Query Match           74.2%;  Score 49;  DB 1;  Length 2464;
  Best Local Similarity  60.0%;  Pred. No. 3.00e+00;
  Matches    6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    463 ISYLTSVSSL 472
          :|||:: |||
Qy      1 VSYLSTASSL 10


RESULT    9
ID   MAPB_HUMAN     STANDARD;      PRT;  2468 AA.
AC   P46821;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   MICROTUBULE-ASSOCIATED PROTEIN 1B [CONTAINS: LIGHT CHAIN LC1].
GN   MAP1B.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95104835.
RA   LIEN L.L., FEENER C., FISCHBACH N., KUNKEL L.M.;
RT   "Cloning of human microtubule-associated protein 1B and the
RT   identification of a related gene on chromosome 15.";
RL   GENOMICS 22:273-280(1994).
CC   -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC       PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC       THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC       TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC       MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC       STABILIZING MICROTUBULES.
CC   -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC       WITH MAP1A AND MAP1B PROTEINS.
CC   -!- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE
CC       KKEE AND KKEI/V, REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER
CC       REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES
CC       BOTH IN VITRO AND IN VIVO.
```

```
CC   -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B.
CC   -!- SIMILARITY: TO NEURAXIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L06237; G473431; -.
DR   MIM; 157129; -.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 6.
DR   PFAM; PF00414; MAP1B_neuraxin; 10.
KW   MICROTUBULES; REPEAT; PHOSPHORYLATION.
FT   CHAIN         ?   2468       MAP1B LC1.
FT   DOMAIN      589    790       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                KKEE AND KKEI/V REPEATS).
FT   DOMAIN     1869   2074       12 X 17 AA TANDEM REPEATS.
FT   REPEAT     1869   1885       1.
FT   REPEAT     1886   1902       2.
FT   REPEAT     1903   1919       3.
FT   REPEAT     1920   1936       4.
FT   REPEAT     1937   1953       5.
FT   REPEAT     1954   1970       6.
FT   REPEAT     1971   1987       7.
FT   REPEAT     1988   2004       8.
FT   REPEAT     2005   2021       9.
FT   REPEAT     2022   2038       10.
FT   REPEAT     2039   2055       11.
FT   REPEAT     2056   2072       12.
SQ   SEQUENCE   2468 AA;  270618 MW;  862C5F69 CRC32;

  Query Match           74.2%;  Score 49;  DB 1;  Length 2468;
  Best Local Similarity  60.0%;  Pred. No. 3.00e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    463 ISYLTSVSSL 472
          :|||:: |||
Qy      1 VSYLSTASSL 10


RESULT   10
ID   MBP_RAT        STANDARD;      PRT;   127 AA.
AC   P02688;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   MYELIN BASIC PROTEIN S (MBP S).
GN   MBP.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87026249.
RA   SCHAICH M., BUDZINSKI R.M., STOFFEL W.;
RT   "Cloned proteolipid protein and myelin basic protein cDNA.
RT   Transcription of the two genes during myelination.";
RL   BIOL. CHEM. HOPPE-SEYLER 367:825-834(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84026484.
RA   ROACH A., BOYLAN K., HORVATH S., PRUSINER S.B., HOOD L.E.;
RT   "Characterization of cloned cDNA representing rat myelin basic
RT   protein: absence of expression in brain of shiverer mutant mice.";
RL   CELL 34:799-806(1983).
RN   [3]
RP   SEQUENCE.
RX   MEDLINE; 75127359.
RA   DUNKLEY P.R., CARNEGIE P.R.;
RT   "Amino acid sequence of the smaller basic protein from rat brain
RT   myelin.";
RL   BIOCHEM. J. 141:243-255(1974).
RN   [4]
RP   SEQUENCE OF 45-85.
RX   MEDLINE; 73180720.
RA   MCFARLIN D.E., BLANK S.E., KIBLER R.F., MCKNEALLY S., SHAPIRA R.;
RT   "Experimental allergic encephalomyelitis in the rat: response to
RT   encephalitogenic proteins and peptides.";
RL   SCIENCE 179:478-480(1973).
CC   -!- FUNCTION: THIS PROTEIN MAY FUNCTION TO MAINTAIN PROPER STRUCTURE
CC       OF MYELIN.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC SIDE OF MYELIN.
CC   -!- ALTERNATIVE PRODUCTS: RATS HAVE TWO MYELIN BASIC PROTEINS. THE
CC       SMALLER ONE, SHOWN HERE, IS MISSING 40 RESIDUES (FOLLOWING RESIDUE
CC       113 OR 114) WITH RESPECT TO THE LARGER ONES FROM OTHER SPECIES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25889; G205322; -.
DR   EMBL; K00512; -; NOT_ANNOTATED_CDS.
DR   PIR; A03142; MBRTS.
DR   PIR; B24351; B24351.
DR   PIR; A21062; A21062.
DR   PROSITE; PS00569; MYELIN_MBP; 1.
KW   MYELIN; STRUCTURAL PROTEIN; ACETYLATION; METHYLATION; PHOSPHORYLATION;
KW   AUTOIMMUNE ENCEPHALOMYELITIS; ALTERNATIVE SPLICING.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION.
FT   MOD_RES     104    104       METHYLATION (MONO-:44% OR DI-:11%).
FT   CONFLICT     46     47       SG -> GS (IN REF. 4).
FT   CONFLICT    124    124       M -> I (IN REF. 2).
SQ   SEQUENCE   127 AA;  14080 MW;  834FEBF5 CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 127;
  Best Local Similarity  66.7%;  Pred. No. 8.55e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db     14 YLATASTMD 22
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT   11
ID   MBP_CAVPO      STANDARD;      PRT;   167 AA.
AC   P25188;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT   01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE   MYELIN BASIC PROTEIN (MBP).
OS   CAVIA PORCELLUS (GUINEA PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; HYSTRICOGNATHI; CAVIIDAE; CAVIA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 84215086.
RA   DEIBLER G.E., MARTENSON R.E., KRUTZSCH H.C., KIES M.W.;
RT   "Sequence of guinea pig myelin basic protein.";
RL   J. NEUROCHEM. 43:100-105(1984).
RN   [2]
RP   SEQUENCE OF 45-87.
RA   SHAPIRA R., MCKNEALLY S.S., CHOU F., KIBLER R.F.;
RT   "Encephalitogenic fragment of myelin basic protein. Amino acid
RT   sequence of bovine, rabbit, guinea pig, monkey, and human fragments.";
RL   J. BIOL. CHEM. 246:4630-4640(1971).
CC   -!- FUNCTION: THIS PROTEIN MAY FUNCTION TO MAINTAIN PROPER STRUCTURE
```

```
CC       OF MYELIN.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC SIDE OF MYELIN.
DR   PIR; A37246; A37246.
DR   PIR; C92087; C92087.
DR   PROSITE; PS00569; MYELIN_MBP; 1.
KW   MYELIN; STRUCTURAL PROTEIN; ACETYLATION; METHYLATION; PHOSPHORYLATION;
KW   AUTOIMMUNE ENCEPHALOMYELITIS.
FT   MOD_RES       1      1       ACETYLATION.
FT   MOD_RES     106    106       METHYLATION (MONO- OR DI-) (BY
FT                                SIMILARITY).
FT   DOMAIN      114    122       INDUCES EXPERIMENTAL AUTOIMMUNE
FT                                ENCEPHALOMYELITIS.
SQ   SEQUENCE   167 AA;  18213 MW;  B31793AB CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 167;
  Best Local Similarity  66.7%;  Pred. No. 8.55e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db     14 YLATASTMD 22
          ||:|||::|
        3 YLSTASSLD 11


RESULT   12
ID   MBP_PANTR      STANDARD;      PRT;   171 AA.
AC   P06906;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   MYELIN BASIC PROTEIN (MBP).
GN   MBP.
OS   PAN TROGLODYTES (CHIMPANZEE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; PAN.
RN   [1]
RP   PRELIMINARY SEQUENCE.
RX   MEDLINE; 76009821.
RA   WESTALL F.C., THOMPSON M., KALTER S.S.;
RT   "The proposed sequence of the encephalitogenic protein from
RT   chimpanzee brain.";
RL   LIFE SCI. 17:219-223(1975).
CC   -!- FUNCTION: THIS PROTEIN MAY FUNCTION TO MAINTAIN PROPER STRUCTURE
CC       OF MYELIN.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC SIDE OF MYELIN.
DR   PIR; A03139; MBCZB.
DR   PROSITE; PS00569; MYELIN_MBP; 1.
KW   MYELIN; STRUCTURAL PROTEIN; ACETYLATION; METHYLATION; PHOSPHORYLATION;
KW   AUTOIMMUNE ENCEPHALOMYELITIS.
FT   MOD_RES       1      1       ACETYLATION.
FT   MOD_RES     107    107       METHYLATION (BY SIMILARITY).
SQ   SEQUENCE   171 AA;  18560 MW;  A616427A CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 171;
  Best Local Similarity  66.7%;  Pred. No. 8.55e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db     14 YLATASTMD 22
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT   13
ID   MBP_MOUSE      STANDARD;      PRT;   194 AA.
AC   P04370;
DT   20-MAR-1987 (REL. 04, CREATED)
DT   01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   MYELIN BASIC PROTEIN (MBP).
GN   MBP.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A. (ALL FOUR FORMS).
RX   MEDLINE; 86079555.
RA   DE FERRA F., ENGH H., HUDSON L., KAMHOLZ J., PUCKETT C., MOLINEAUX S.,
RA   LAZZARINI R.A.;
RT   "Alternative splicing accounts for the four forms of myelin basic
RT   protein.";
RL   CELL 43:721-727(1985).
RN   [2]
RP   SEQUENCE FROM N.A. (18.5 KD FORM).
RX   MEDLINE; 85254913.
RA   TAKAHASHI N., ROACH A., TEPLOW D.B., PRUSINER S.B., HOOD L.;
RT   "Cloning and characterization of the myelin basic protein gene from
RT   mouse: one gene can encode both 14 kd and 18.5 kd MBPs by alternate
RT   use of exons.";
RL   CELL 42:139-148(1985).
RN   [3]
RP   SEQUENCE FROM N.A. (17 KD FORM).
RX   MEDLINE; 87118269.
RA   NEWMAN S., KITAMURA K., CAMPAGNONI A.T.;
RT   "Identification of a cDNA coding for a fifth form of myelin basic
RT   protein in mouse.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 84:886-890(1987).
RN   [4]
RP   SEQUENCE OF 1-23 FROM N.A.
RX   MEDLINE; 89252919.
RA   MIURA M., TAMURA T.A., AOYAMA A., MIKOSHIBA K.;
RT   "The promoter elements of the mouse myelin basic protein gene
RT   function efficiently in NG108-15 neuronal/glial cells.";
RL   GENE 75:31-38(1989).
CC   -!- FUNCTION: THIS PROTEIN MAY FUNCTION TO MAINTAIN PROPER STRUCTURE
CC       OF MYELIN.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC SIDE OF MYELIN.
CC   -!- ALTERNATIVE PRODUCTS: MOUSE HAS FOUR FORMS OF MBP, 21.5KD, 18.5KD,
CC       17KD, AND 14KD, PRESENT IN RELATIVE AMOUNTS OF 1:10:3.5:35. THOSE
CC       4 FORMS ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- THE SEQUENCE SHOWN IS THAT OF THE 21.5KD FORM.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11533; G387414; -.
DR   EMBL; M11291; G387414; JOINED.
DR   EMBL; M11529; G387414; JOINED.
DR   EMBL; M11530; G387414; JOINED.
DR   EMBL; M11531; G387414; JOINED.
DR   EMBL; M11532; G387414; JOINED.
DR   EMBL; M11533; G387415; -.
DR   EMBL; M11291; G387415; JOINED.
DR   EMBL; M11529; G387415; JOINED.
DR   EMBL; M11530; G387415; JOINED.
DR   EMBL; M11531; G387415; JOINED.
DR   EMBL; L00404; G387416; -.
DR   EMBL; L00398; G387416; JOINED.
DR   EMBL; L00400; G387416; JOINED.
DR   EMBL; L00401; G387416; JOINED.
DR   EMBL; L00402; G387416; JOINED.
DR   EMBL; L00404; G387417; -.
DR   EMBL; L00398; G387417; JOINED.
DR   EMBL; L00399; G387417; JOINED.
DR   EMBL; L00400; G387417; JOINED.
DR   EMBL; L00401; G387417; JOINED.
DR   EMBL; L00402; G387417; JOINED.
DR   EMBL; L00404; G387418; -.
DR   EMBL; L00398; G387418; JOINED.
DR   EMBL; L00400; G387418; JOINED.
DR   EMBL; L00401; G387418; JOINED.
DR   EMBL; L00402; G387418; JOINED.
DR   EMBL; L00403; G387418; JOINED.
```

```
DR   EMBL; L00404; G387419; -.
DR   EMBL; L00398; G387419; JOINED.
DR   EMBL; L00399; G387419; JOINED.
DR   EMBL; L00400; G387419; JOINED.
DR   EMBL; L00401; G387419; JOINED.
DR   EMBL; L00402; G387419; JOINED.
DR   EMBL; L00403; G387419; JOINED.
DR   EMBL; M15060; G199049; -.
DR   EMBL; M24410; G554196; -.
DR   PIR; A24772; MBMSB.
DR   PIR; A26591; A26591.
DR   PIR; B26591; B26591.
DR   MGD; MGI:96925; MBP.
DR   PROSITE; PS00569; MYELIN_MBP; 1.
KW   MYELIN; STRUCTURAL PROTEIN; ACETYLATION; METHYLATION; PHOSPHORYLATION;
KW   AUTOIMMUNE ENCEPHALOMYELITIS; ALTERNATIVE SPLICING.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   MOD_RES     130    130       METHYLATION (BY SIMILARITY).
FT   VARSPLIC     57     82       MISSING (IN 18.5 KD FORM AND 14 KD FORM).
FT   VARSPLIC    140    180       MISSING (IN 17 KD FORM AND 14 KD FORM).
     SEQUENCE   194 AA;  21371 MW;  5B6D9A74 CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 194;
  Best Local Similarity  66.7%;  Pred. No. 8.55e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db     12 YLATASTMD 20
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT   14
ID   MBP_HUMAN      STANDARD;      PRT;   196 AA.
AC   P02686;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT   01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE   MYELIN BASIC PROTEIN (MBP).
GN   MBP.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89302693.
RA   STREICHER R., STOFFEL W.;
RT   "The organization of the human myelin basic protein gene. Comparison
     with the mouse gene.";
     BIOL. CHEM. HOPPE-SEYLER 370:503-510(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87311781.
RA   ROTH H.J., KRONQUIST K.E., DE ROSBO N., CRANDALL B.F.,
RA   CAMPAGNONI A.T.;
RT   "Evidence for the expression of four myelin basic protein variants in
RT   the developing human spinal cord through cDNA cloning.";
RL   J. NEUROSCI. RES. 17:321-328(1987).
RN   [3]
RP   SEQUENCE OF 1-58; 85-131 AND 143-196 FROM N.A.
RX   MEDLINE; 86308101.
RA   ROTH H.J., KRONQUIST K., PRETORIUS P.J., CRANDALL B.F.,
RA   CAMPAGNONI A.T.;
RT   "Isolation and characterization of a cDNA coding for a novel human
RT   17.3K myelin basic protein (MBP) variant.";
RL   J. NEUROSCI. RES. 16:227-238(1986).
RN   [4]
RP   SEQUENCE OF 1-58 AND 85-196 FROM N.A.
RX   MEDLINE; 86259714.
RA   KAMHOLZ J., DE FERRA F., PUCKETT C., LAZZARINI R.A.;
RT   "Identification of three forms of human myelin basic protein by cDNA
RT   cloning.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 83:4962-4966(1986).
RN   [5]
RP   SEQUENCE OF 1-58 AND 85-196.
RX   MEDLINE; 72066400.
RA   CARNEGIE P.R.;
RT   "Amino acid sequence of the encephalitogenic basic protein from human
RT   myelin.";
RL   BIOCHEM. J. 123:57-67(1971).
RN   [6]
RP   SEQUENCE OF 45-58 AND 85-114, REVISIONS.
RA   SHAPIRA R., MCKNEALLY S.S., CHOU F., KIBLER R.F.;
RT   "Encephalitogenic fragment of myelin basic protein. Amino acid
RT   sequence of bovine, rabbit, guinea pig, monkey, and human fragments.";
RL   J. BIOL. CHEM. 246:4630-4640(1971).
RN   [7]
RP   SEQUENCE OF 1-58 FROM N.A.
RX   MEDLINE; 90152679.
RA   BOYLAN K.B., AYRES T.M., POPKO B., TAKAHASHI N., HOOD L.E.,
RA   PRUSINER S.B.;
RT   "Repetitive DNA (TGGA)n 5' to the human myelin basic protein gene: a
RT   new form of oligonucleotide repetitive sequence showing length
RT   polymorphism.";
RL   GENOMICS 6:16-22(1990).
RN   [8]
RP   METHYLATION.
RX   MEDLINE; 72066401.
RA   BALDWIN G.S., CARNEGIE P.R.;
RT   "Isolation and partial characterization of methylated arginines from
RT   the encephalitogenic basic protein of myelin.";
RL   BIOCHEM. J. 123:69-74(1971).
CC   -!- FUNCTION: THIS PROTEIN MAY FUNCTION TO MAINTAIN PROPER STRUCTURE
CC       OF MYELIN.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC SIDE OF MYELIN.
CC   -!- ALTERNATIVE PRODUCTS: HUMANS HAVE THREE FORMS OF MBP, 21.5K, 18.5K
CC       (THE MOST ABUNDANT), AND 17.2K. THE THREE FORMS ARE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M13577; G307160; -.
DR   EMBL; M30516; G307161; -.
DR   EMBL; M30515; G307162; -.
DR   EMBL; M30047; G307159; -.
DR   EMBL; X17286; E221974; -.
DR   EMBL; X17287; E221974; JOINED.
DR   EMBL; X17290; E221974; JOINED.
DR   EMBL; X17288; E221974; JOINED.
DR   EMBL; X17369; E221974; JOINED.
DR   EMBL; X17289; E221974; JOINED.
DR   EMBL; M63599; G187403; -.
DR   PIR; A24153; MBHUB.
DR   PIR; S10482; S10482.
DR   MIM; 159430; -.
DR   PROSITE; PS00569; MYELIN_MBP; 1.
KW   MYELIN; STRUCTURAL PROTEIN; ACETYLATION; METHYLATION; PHOSPHORYLATION;
KW   AUTOIMMUNE ENCEPHALOMYELITIS; ALTERNATIVE SPLICING.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION.
FT   MOD_RES     133    133       METHYLATION (MONO-:6% OR DI-:60%).
FT   DOMAIN       45    114       INDUCES EXPERIMENTAL AUTOIMMUNE
FT                                ENCEPHALOMYELITIS IN MONKEYS, RABBITS,
FT                                AND RATS BUT NOT IN GUINEA PIGS.
FT   DOMAIN      138    148       INDUCES EXPERIMENTAL AUTOIMMUNE
FT                                ENCEPHALOMYELITIS.
FT   VARSPLIC     59     84       MISSING (IN 18.5 KD FORM AND 17.2 KD
FT                                FORM).
FT   VARSPLIC    132    142       MISSING (IN 17.2 KD FORM).
SQ   SEQUENCE   196 AA;  21362 MW;  0A8864EE CRC32;
```

```
  Query Match           71.2%;  Score 47;  DB 1;  Length 196;
  Best Local Similarity  66.7%;  Pred. No. 8.55e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db      14 YLATASTMD 22
           ||:|||::|
Qy       3 YLSTASSLD 11


RESULT   15
ID   RNPH_MYCTU     STANDARD;      PRT;   259 AA.
AC   Q10628;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PROBABLE RIBONUCLEASE PH (EC 2.7.7.56) (RNASE PH) (TRNA
DE   NUCLEOTIDYLTRANSFERASE).
GN   RPH OR MTCY130.25 OR MTCY02B10.04.
OS   MYCOBACTERIUM TUBERCULOSIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: RNASE PH IS A PHOSPHOROLYTIC EXORIBONUCLEASE THAT
CC       REMOVES NUCLEOTIDE RESIDUES FOLLOWING THE -CCA TERMINUS OF TRNA
CC       AND ADDS NUCLEOTIDES TO THE ENDS OF RNA MOLECULES BY USING
CC       NUCLEOSIDE DIPHOSPHATES AS SUBSTRATES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: TRNA(N+1) + ORTHOPHOSPHATE = TRNA(N) +
CC       A NUCLEOSIDE DIPHOSPHATE.
CC   -!- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73902; E245034; -.
DR   EMBL; Z75555; E250346; -.
DR   PROSITE; PS01277; RIBONUCLEASE_PH; 1.
DR   PFAM; PF01138; RNase_PH; 1.
KW   TRANSFERASE; NUCLEOTIDYLTRANSFERASE; TRNA PROCESSING.
SQ   SEQUENCE   259 AA;  27351 MW;  89F9EFE6 CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 259;
  Best Local Similarity  60.0%;  Pred. No. 8.55e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     143 VTYLSAAGKL 152
           |:|||:|: |
Qy       1 VSYLSTASSL 10


Search completed: Thu Sep  2 11:21:33 1999
Job time : 8 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 11:20:54 1999; MasPar time 3.25 Seconds
147.929 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-4
Description: (1-12) from US08599226.pep
Perfect Score: 66
Sequence: 1 VSYLSTASSLDX 12

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.694; Variance 27.579; scale 0.859

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	52	78.8	547	2	E29504	mercury(II) reductase	2.07e+00
2	51	77.3	236	2	S60941	probable membrane pro	3.40e+00
3	50	75.8	293	2	D70108	conserved hypothetica	5.54e+00
4	49	74.2	132	2	H70759	hypothetical protein	8.98e+00
5	49	74.2	300	2	A71645	protein p34 (p34) RP8	8.98e+00
6	49	74.2	471	2	S08325	flavonol 3-O-glucosyl	8.98e+00
7	49	74.2	471	2	S01052	flavonol 3-O-glucosyl	8.98e+00
8	49	74.2	471	2	S01037	flavonol 3-O-glucosyl	8.98e+00
9	49	74.2	599	2	G71481	probable phosphoenolp	8.98e+00
10	49	74.2	2364	2	A56577	microtubule-associate	8.98e+00
11	49	74.2	2464	1	QRMSP1	microtubule-associate	8.98e+00
12	48	72.7	292	2	T03085	ribonuclease homolog	1.45e+01
13	48	72.7	832	2	S71788	P/CAF protein - human	1.45e+01
14	47	71.2	128	1	MBRTS	myelin basic protein	2.31e+01
15	47	71.2	167	2	A37246	myelin basic protein	2.31e+01
16	47	71.2	171	1	MBCZB	myelin basic protein	2.31e+01
17	47	71.2	197	1	MBHUB	myelin basic protein	2.31e+01
18	47	71.2	239	2	S25204	srmX protein - Strept	2.31e+01
19	47	71.2	259	2	C70739	probable rphA protein	2.31e+01
20	47	71.2	259	1	S72788	tRNA nucleotidyltrans	2.31e+01
21	47	71.2	328	1	MBMSB	golli-myelin basic pr	2.31e+01
22	47	71.2	1107	2	E65226	hypothetical 123.8 kD	2.31e+01
23	46	69.7	133	2	C71485	hypothetical protein	3.66e+01
24	46	69.7	474	2	S30168	mercury(II) reductase	3.66e+01
25	46	69.7	485	2	B37855	glucose-6-phosphate 1	3.66e+01
26	46	69.7	500	2	B64680	anthranilate synthase	3.66e+01
27	46	69.7	500	2	G71836	anthranilate synthase	3.66e+01
28	46	69.7	554	2	JU0188	membrane protein 4.1	3.66e+01
29	46	69.7	564	2	S57124	CTP synthase (EC 6.3.	3.66e+01
30	46	69.7	687	2	A41905	ferric vibriobactin r	3.66e+01
31	46	69.7	761	2	C64813	ybhJ protein - Escher	3.66e+01
32	45	68.2	87	3	D42603	ORF 3' to arbB - Erwi	5.75e+01
33	45	68.2	261	2	A71192	probable glucose 1-de	5.75e+01
34	45	68.2	333	2	S74406	hypothetical protein	5.75e+01
35	45	68.2	416	2	H64984	hypothetical 43.4 kD	5.75e+01
36	45	68.2	471	2	C71439	hypothetical protein	5.75e+01
37	45	68.2	505	2	S76229	hypothetical protein	5.75e+01
38	45	68.2	566	2	G33157	hemagglutinin precurs	5.75e+01
39	45	68.2	608	2	T02684	DNA-binding protein C	5.75e+01
40	45	68.2	619	2	A56519	CDC16 protein - human	5.75e+01
41	45	68.2	631	2	E32227	mercury(II) reductase	5.75e+01
42	45	68.2	633	2	S49788	probable membrane pro	5.75e+01
43	45	68.2	650	2	F70974	probable acrA1 protei	5.75e+01
44	45	68.2	1369	2	JC4860	protein-tyrosine kina	5.75e+01
45	45	68.2	1545	2	S48824	probable membrane pro	5.75e+01

ALIGNMENTS

RESULT 1
ENTRY E29504 #type complete
TITLE mercury(II) reductase (EC 1.16.1.1) - Staphylococcus aureus plasmid pI258
ALTERNATE_NAMES Hg(II) reductase; mercuric reductase
ORGANISM #formal_name Staphylococcus aureus
DATE 21-May-1988 #sequence_revision 21-May-1988 #text_change 05-Dec-1998
ACCESSIONS E29504
REFERENCE A94170
#authors Laddaga, R.A.; Chu, L.; Misra, T.K.; Silver, S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:5106-5110
#title Nucleotide sequence and expression of the mercurial resistance operon from Staphylococcus aureus plasmid pI258.
#cross-references MUID:87260937
#accession E29504
##molecule_type DNA
##residues 1-547 ##label LAD
GENETICS
#gene merA
#genome plasmid
CLASSIFICATION #superfamily Bacillus mercury(II) reductase; dihydrolipoamide dehydrogenase homology; heavy-metal-associated homology
KEYWORDS disulfide bond; FAD; oxidoreductase
FEATURE
10-38 #domain heavy-metal-associated homology #label HMA\
89-528 #domain dihydrolipoamide dehydrogenase homology #label DLD
SUMMARY #length 547 #molecular-weight 58565 #checksum 1158

Query Match 78.8%; Score 52; DB 2; Length 547;
Best Local Similarity 45.5%; Pred. No. 2.07e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```
Db    235 VDYLTSTSALE 245
          | ||:::|:|:
Qy      1 VSYLSTASSLD 11
```

RESULT 2
ENTRY S60941 #type complete
TITLE probable membrane protein YOR214c - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein O5022; hypothetical protein YOR50-4
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 05-Jun-1998

```
ACCESSIONS       S60941; S67107; S71716
REFERENCE        S60938
   #authors      Galisson, F.; Dujon, B.
   #submission   submitted to the EMBL Data Library, October 1995
   #description  Sequence and analysis of a 33 kb fragment from the right arm
                   of chromosome XV of the yeast Saccharomyces cerevisiae.
   #accession    S60941
      ##molecule_type DNA
      ##residues      1-236 ##label GAL
      ##cross-references EMBL:X92441; NID:g1050762; PID:g1050766
REFERENCE        S67104
   #authors      Boyer, J.; Fairhead, C.; Gaillon, L.; Galisson, F.; Michaux,
                   G.; Thierry, A.; Dujon, B.
   #submission   submitted to the Protein Sequence Database, July 1996
   #accession    S67107
      ##molecule_type DNA
      ##residues      1-236 ##label BOY
      ##cross-references EMBL:Z75122; NID:g1420498; PID:e252081; PID:g1420499;
                        MIPS:YOR214c
      ##experimental_source strain S288C
REFERENCE        S71713
   #authors      Galisson, F.; Dujon, B.
   #journal      Yeast (1996) 12:877-885
   #title        Sequence and analysis of a 33 kb fragment from the right arm
                   of chromosome XV of the yeast Saccharomyces cerevisiae.
   #cross-references MUID:96437977
   #accession    S71716
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-236 ##label GAW
      ##cross-references EMBL:X92441; NID:g1050762; PID:g1050766
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, October 1995
GENETICS
   #map_position 15R
KEYWORDS         transmembrane protein
FEATURE
   4-20              #domain transmembrane #status predicted #label TM1\
   217-233           #domain transmembrane #status predicted #label TM2
SUMMARY          #length 236  #molecular-weight 26156  #checksum 8999

  Query Match           77.3%;  Score 51;  DB 2;  Length 236;
  Best Local Similarity  60.0%;  Pred. No. 3.40e+00;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     175 ISYLSSTTSL 184
           :||||:::||
Qy       1 VSYLSTASSL 10


RESULT    3
ENTRY            D70108     #type complete
TITLE            conserved hypothetical protein BB0068 - Lyme disease
                   spirochete
ORGANISM         #formal_name Borrelia burgdorferi #common_name Lyme disease
                   spirochete
DATE             13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                   05-Jun-1998
ACCESSIONS       D70108
REFERENCE        A70100
   #authors      Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
                   Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
                   Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
                   J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
                   Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
                   Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
                   J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
                   Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
                   K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
   #journal      Nature (1997) 390:580-586
   #title        Genomic sequence of a Lyme disease spirochaete, Borrelia
                   burgdorferi.
   #cross-references MUID:98065943
   #accession    D70108
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-293 ##label KLE
      ##cross-references GB:AE001120; GB:AE000783; NID:g2687951; PID:g2687956;
                        TIGR:BB0068
      ##experimental_source strain B31
SUMMARY          #length 293  #molecular-weight 33278  #checksum 5223

  Query Match           75.8%;  Score 50;  DB 2;  Length 293;
  Best Local Similarity  60.0%;  Pred. No. 5.54e+00;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     198 AYLSTPNSLE 207
           :||||::||:
Qy       2 SYLSTASSLD 11


RESULT    4
ENTRY            H70759     #type complete
TITLE            hypothetical protein Rv2010 - Mycobacterium tuberculosis
                   (strain H37RV)
ORGANISM         #formal_name Mycobacterium tuberculosis
DATE             17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                   17-Jul-1998
ACCESSIONS       H70759
REFERENCE        A70500
   #authors      Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                   C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                   III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                   Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                   Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                   Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                   Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                   Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                   Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                   Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal      Nature (1998) 393:537-544
   #title        Deciphering the biology of Mycobacterium tuberculosis from
                   the complete genome sequence.
   #cross-references MUID:98295987
   #accession    H70759
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-132 ##label COL
      ##cross-references GB:Z74025; GB:AL123456; NID:g3261586; PID:e248788;
                        PID:g1403443
      ##experimental_source strain H37Rv
GENETICS
   #gene         Rv2010
SUMMARY          #length 132  #molecular-weight 14731  #checksum 1326

  Query Match           74.2%;  Score 49;  DB 2;  Length 132;
  Best Local Similarity  60.0%;  Pred. No. 8.98e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db       9 IAYLSTSESL 18
           ::||||: ||
Qy       1 VSYLSTASSL 10


RESULT    5
ENTRY            A71645     #type complete
TITLE            protein p34 (p34) RP832 - Rickettsia prowazekii
ORGANISM         #formal_name Rickettsia prowazekii
DATE             21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
                   21-Nov-1998
ACCESSIONS       A71645
REFERENCE        A71630
   #authors      Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
                   Sicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
```

```
                  Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
                  C.G.
   #journal       Nature (1998) 396:133-140
   #title         The genome sequence of Rickettsia prowazekii and the origin
                    of mitochondria.
   #accession     A71645
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-300 ##label AND
      ##cross-references GB:AJ235273; GB:AJ235269; NID:g3861237; PID:e1343102;
                         PID:g3861358
      ##experimental_source strain Madrid E
GENETICS
   #gene          p34; RP832
SUMMARY           #length 300  #molecular-weight 34582  #checksum 8509

  Query Match           74.2%;  Score 49;  DB 2;  Length 300;
  Best Local Similarity  60.0%;  Pred. No. 8.98e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db      12 VSYLSVTTAL 21
           ||||| :::|
Qy       1 VSYLSTASSL 10


RESULT    6
ENTRY             S08325     #type complete
TITLE             flavonol 3-O-glucosyltransferase (EC 2.4.1.91) (allele
                    BzMcC2) - maize
ALTERNATE_NAMES   UDPglucose flavonoid glucosyl-transferase
ORGANISM          #formal_name Zea mays #common_name maize
DATE              30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
                    08-Sep-1997
ACCESSIONS        S08325
REFERENCE         S08324
   #authors       Furtek, D.; Schiefelbein, J.W.; Johnston, F.; Nelson Jr.,
                    O.E.
   #journal       Plant Mol. Biol. (1988) 11:473-481
   #title         Sequence comparisons of three wild-type Bronze-1 alleles from
                    Zea mays.
   #accession     S08325
      ##status         translation not shown
      ##molecule_type DNA
      ##residues       1-471 ##label FUR
      ##cross-references EMBL:X13501; NID:g22361; PID:g295854
GENETICS
   #gene          Bz1
   #introns       175/1
CLASSIFICATION    #superfamily flavonol O3-glucosyltransferase
KEYWORDS          glycosyltransferase; hexosyltransferase
SUMMARY           #length 471  #molecular-weight 48621  #checksum 7439

  Query Match           74.2%;  Score 49;  DB 2;  Length 471;
  Best Local Similarity  80.0%;  Pred. No. 8.98e+00;
  Matches     8;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      46 LSFLSTASSL 55
           :|:|||||||
Qy       1 VSYLSTASSL 10


RESULT    7
ENTRY             S01052     #type complete
TITLE             flavonol 3-O-glucosyltransferase (EC 2.4.1.91) (allele
                    Bz-McC) - maize
ALTERNATE_NAMES   UDPglucose flavonoid glucosyltransferase
ORGANISM          #formal_name Zea mays #common_name maize
DATE              31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
                    08-Sep-1997
ACCESSIONS        S01052; S08324
REFERENCE         S01037
   #authors       Ralston, E.J.; English, J.J.; Dooner, H.K.
   #journal       Genetics (1988) 119:185-197
   #title         Sequence of three bronze alleles of maize and correlation
                    with the genetic fine structure.
   #cross-references MUID:88284304
   #accession     S01052
      ##status         translation not shown
      ##molecule_type DNA
      ##residues       1-471 ##label RAL
      ##cross-references EMBL:X07940; NID:g22204; PID:g22205
REFERENCE         S08324
   #authors       Furtek, D.; Schiefelbein, J.W.; Johnston, F.; Nelson Jr.,
                    O.E.
   #journal       Plant Mol. Biol. (1988) 11:473-481
   #title         Sequence comparisons of three wild-type Bronze-1 alleles from
                    Zea mays.
   #accession     S08324
      ##status         translation not shown
      ##molecule_type DNA
      ##residues       1-471 ##label FUR
      ##cross-references EMBL:X13500; NID:g22364; PID:g1030071
GENETICS
   #gene          Bz1
   #map_position 9
   #introns       175/1
CLASSIFICATION    #superfamily flavonol O3-glucosyltransferase
KEYWORDS          glycosyltransferase; hexosyltransferase
SUMMARY           #length 471  #molecular-weight 48769  #checksum 6660

  Query Match           74.2%;  Score 49;  DB 2;  Length 471;
  Best Local Similarity  80.0%;  Pred. No. 8.98e+00;
  Matches     8;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      46 LSFLSTASSL 55
           :|:|||||||
Qy       1 VSYLSTASSL 10


RESULT    8
ENTRY             S01037     #type complete
TITLE             flavonol 3-O-glucosyltransferase (EC 2.4.1.91) (allele
                    Bz-W22) - maize
ALTERNATE_NAMES   UDPglucose flavonoid glucosyltransferase
ORGANISM          #formal_name Zea mays #common_name maize
DATE              31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
                    08-Sep-1997
ACCESSIONS        S01037; S08326
REFERENCE         S01037
   #authors       Ralston, E.J.; English, J.J.; Dooner, H.K.
   #journal       Genetics (1988) 119:185-197
   #title         Sequence of three bronze alleles of maize and correlation
                    with the genetic fine structure.
   #cross-references MUID:88284304
   #accession     S01037
      ##status         translation not shown
      ##molecule_type DNA
      ##residues       1-471 ##label RAL
      ##cross-references EMBL:X07937; NID:g22209; PID:g22210
REFERENCE         S08324
   #authors       Furtek, D.; Schiefelbein, J.W.; Johnston, F.; Nelson Jr.,
                    O.E.
   #journal       Plant Mol. Biol. (1988) 11:473-481
   #title         Sequence comparisons of three wild-type Bronze-1 alleles from
                    Zea mays.
   #accession     S08326
      ##status         translation not shown
      ##molecule_type mRNA
      ##residues       1-471 ##label FUR
      ##cross-references EMBL:X13502; NID:g22505; PID:g22506
GENETICS
   #gene          Bz1
   #map_position 9
   #introns       175/1
CLASSIFICATION    #superfamily flavonol O3-glucosyltransferase
```

```
KEYWORDS          anthocyanin biosynthesis; glycosyltransferase;
                    hexosyltransferase
SUMMARY           #length 471  #molecular-weight 48673  #checksum 7590

  Query Match           74.2%;  Score 49;  DB 2;  Length 471;
  Best Local Similarity  80.0%;  Pred. No. 8.98e+00;
  Matches     8;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     46 LSFLSTASSL 55
          :|:|||||||
Qy      1 VSYLSTASSL 10


RESULT    9
ENTRY             G71481     #type complete
TITLE             probable phosphoenolpyruvate carboxykinase - Chlamydia
                    trachomatis (serotype D, strain UW3/Cx)
ORGANISM          #formal_name Chlamydia trachomatis
DATE              13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
                    21-Nov-1998
ACCESSIONS        G71481
REFERENCE         A71570
   #authors       Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
                    R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
                    R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
   #journal       Science (1998) 282:754-759
   #title         Genome sequence of an obligate intracellular pathogen of
                    humans: Chlamydia trachomatis.
   #cross-references MUID:99000809
   #accession     G71481
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-599 ##label ARN
      ##cross-references GB:AE001341; GB:AE001273; NID:g3329156; PID:g3329165
      ##experimental_source serotype D, strain UW-3/Cx
GENETICS
   #gene          pckA
CLASSIFICATION    #superfamily phosphoenolpyruvate carboxykinase (GTP)
SUMMARY           #length 599  #molecular-weight 66244  #checksum 5490

  Query Match           74.2%;  Score 49;  DB 2;  Length 599;
  Best Local Similarity  54.5%;  Pred. No. 8.98e+00;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    533 VGYLPTAEGLN 543
          |:||:|| :|:
Qy      1 VSYLSTASSLD 11


RESULT   10
ENTRY             A56577     #type fragment
TITLE             microtubule-associated protein MAP 1B - rat (fragment)
ORGANISM          #formal_name Rattus norvegicus #common_name Norway rat
DATE              21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change
                    16-Feb-1997
ACCESSIONS        A56577
REFERENCE         A56577
   #authors       Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.
   #journal       Eur. J. Cell Biol. (1992) 57:66-74
   #title         Identification of two distinct microtubule binding domains on
                    recombinant rat MAP 1B.
   #cross-references MUID:92347374
   #accession     A56577
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-2364 ##label ZAU
      ##cross-references GB:X60550
      ##experimental_source brain
      ##note           nucleotide sequence not given; conceptual translation
                        not complete
CLASSIFICATION    #superfamily microtubule-associated protein MAP1B
SUMMARY           #length 2364  #checksum 9159

  Query Match           74.2%;  Score 49;  DB 2;  Length 2364;
  Best Local Similarity  60.0%;  Pred. No. 8.98e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    367 ISYLASVSSL 376
          :|||:: |||
Qy      1 VSYLSTASSL 10


RESULT   11
ENTRY             QRMSP1     #type complete
TITLE             microtubule-associated protein MAP1B - mouse
ALTERNATE_NAMES   microtubule-associated protein MAP1(X);
                    microtubule-associated protein MAP1.2;
                    microtubule-associated protein MAP5
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
                    05-Sep-1997
ACCESSIONS        S07549; A33645
REFERENCE         A33645
   #authors       Noble, M.; Lewis, S.A.; Cowan, N.J.
   #journal       J. Cell Biol. (1989) 109:3367-3376
   #title         The microtubule binding domain of microtubule-associated
                    protein MAP1B contains a repeated sequence motif unrelated
                    to that of MAP2 and tau.
   #cross-references MUID:90094539
   #accession     S07549
      ##molecule_type mRNA
      ##residues       1-2464 ##label NOB
      ##cross-references EMBL:X51396; NID:g52999; PID:g53000
CLASSIFICATION    #superfamily microtubule-associated protein MAP1B
KEYWORDS          microtubule binding; phosphoprotein; tandem repeat
FEATURE
   589-786              #domain microtubule binding #status experimental #label
                          MTB\
   589-592,639-642,
   649-652,655-658,
   660-663,668-671,
   674-677,679-682,
   683-686,687-690,
   691-694,695-698,
   699-702,708-711,
   712-715,716-719,
   720-723,727-730,
   758-761,764-767,
   783-786              #region 4-residue repeats (K/R-K-E/D-X)\
   1861-2064            #region 17-residue repeats\
   91,116,351,888,
   1124,1153,1168,
   1208,1662,1877,
   1918,2003,2030,
   2054,2083            #binding_site phosphate (Ser) (covalent) #status
                          predicted\
   147,969,1336,1562,
   1563,1702,1708,
   1990,2057,2063,
   2419                 #binding_site phosphate (Thr) (covalent) #status
                          predicted\
   1953                 #binding_site phosphate (Tyr) (covalent) #status
                          predicted
SUMMARY           #length 2464  #molecular-weight 270408  #checksum 2525

  Query Match           74.2%;  Score 49;  DB 1;  Length 2464;
  Best Local Similarity  60.0%;  Pred. No. 8.98e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    463 ISYLTSVSSL 472
          :|||:: |||
Qy      1 VSYLSTASSL 10


RESULT   12
ENTRY             T03085     #type complete
```

```
TITLE           ribonuclease homolog 068R - Chilo iridescent virus
ORGANISM        #formal_name Chilo iridescent virus
DATE            24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
                  24-Mar-1999
ACCESSIONS      T03085
REFERENCE       Z14834
   #authors     Bahr, U.; Tidona, C.A.; Darai, G.
   #journal     Virus Genes (1997) 15:235-245
   #title       The DNA sequence of Chilo iridescent virus between the genome
                  coordinates 0.101 and 0.391; similarities in coding
                  strategy between insect and vertebrate iridoviruses.
   #accession   T03085
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-292 ##label BAH
      ##cross-references EMBL:AF003534; NID:g2738385; PID:g2738432
SUMMARY         #length 292  #molecular-weight 33664  #checksum 532

  Query Match           72.7%;  Score 48;  DB 2;  Length 292;
  Best Local Similarity  45.5%;  Pred. No. 1.45e+01;
  Matches    5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     39 IPYLTTPESLN 49
          ::||:|: ||:
Qy      1 VSYLSTASSLD 11


RESULT   13
ENTRY           S71788     #type complete
TITLE           P/CAF protein - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change
                  02-Jul-1998
ACCESSIONS      S71788
REFERENCE       S71788
   #authors     Yang, X.J.; Ogryzko, V.V.; Nishikawa, J.; Howard, B.H.;
                  Nakatani, Y.
   #journal     Nature (1996) 382:319-324
   #title       A p300/CBP-associated factor that competes with the
                  adenoviral oncoprotein E1A.
   #cross-references MUID:96300317
   #accession   S71788
      ##status        preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-832 ##label YAN
      ##cross-references EMBL:U57317; NID:g1491936; PID:g1491937
CLASSIFICATION  #superfamily bromodomain homology
FEATURE
   748-803           #domain bromodomain homology #label BRO1
MMARY           #length 832  #molecular-weight 92926  #checksum 8709

  Query Match           72.7%;  Score 48;  DB 2;  Length 832;
  Best Local Similarity  63.6%;  Pred. No. 1.45e+01;
  Matches    7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    394 ISYNSTSSSLE 404
          :|| ||:|||:
Qy      1 VSYLSTASSLD 11


RESULT   14
ENTRY           MBRTS      #type complete
TITLE           myelin basic protein S - rat
ALTERNATE_NAMES small myelin basic protein
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            24-Apr-1984 #sequence_revision 08-Feb-1996 #text_change
                  05-Sep-1997
ACCESSIONS      B24351; A90275; A94243; A21062; A03142
REFERENCE       A24351
   #authors     Schaich, M.; Budzinski, R.M.; Stoffel, W.
   #journal     Biol. Chem. Hoppe-Seyler (1986) 367:825-834
   #title       Cloned proteolipid protein and myelin basic protein cDNA.
                  Transcription of the two genes during myelination.
   #cross-references MUID:87026249
   #accession   B24351
      ##molecule_type mRNA
      ##residues      1-128 ##label SCH
      ##cross-references EMBL:M25889; NID:g205321; PID:g205322
REFERENCE       A90275
   #authors     Dunkley, P.R.; Carnegie, P.R.
   #journal     Biochem. J. (1974) 141:243-255
   #title       Amino acid sequence of the smaller basic protein from rat
                  brain myelin.
   #cross-references MUID:75127359
   #accession   A90275
      ##molecule_type protein
      ##residues      2-128 ##label DUN
      ##note          at position 105, arginine, monomethylarginine, and
                       dimethylarginine occur in the ratio 4:4:1
      ##note          rats have two myelin basic proteins; the smaller one,
                       shown above, is missing 40 residues (following residue
                       114 or 115) with respect to the larger ones from other
                       species
REFERENCE       A94243
   #authors     McFarlin, D.E.; Blank, S.E.; Kibler, R.F.; McKneally, S.;
                  Shapira, R.
   #journal     Science (1973) 179:478-480
   #title       Experimental allergic encephalomyelitis in the rat: response
                  to encephalitogenic proteins and peptides.
   #cross-references MUID:73180720
   #accession   A94243
      ##molecule_type protein
      ##residues      46-86 ##label MCF
      ##note          the sequence reported for this encephalitogenic peptide
                       differs from that shown by a transposition of residues
                       47 and 48; two other differences are printing errors
REFERENCE       A21062
   #authors     Roach, A.; Boylan, K.; Horvath, S.; Prusiner, S.B.; Hood,
                  L.E.
   #journal     Cell (1983) 34:799-806
   #title       Characterization of cloned cDNA representing rat myelin basic
                  protein: absence of expression in brain of shiverer mutant
                  mice.
   #cross-references MUID:84026484
   #accession   A21062
      ##molecule_type mRNA
      ##residues      1-124,'I',126-128 ##label ROA
      ##experimental_source strain Sprague-Dawley
CLASSIFICATION  #superfamily myelin basic protein
KEYWORDS        alternative splicing; blocked amino end; experimental
                  autoimmune encephalomyelitis; methylated amino acid; myelin

FEATURE
   2-128             #product myelin basic protein S #status experimental
                       #label MAT\
   2                 #modified_site blocked amino end (Ala) (in mature form)
                       (probably acetylated) #status experimental\
   105               #modified_site omega-N-methylarginine or omega-N,
                       omega-N'-dimethylarginine (Arg) (partial) #status
                       experimental
SUMMARY         #length 128  #molecular-weight 14211  #checksum 2812

  Query Match           71.2%;  Score 47;  DB 1;  Length 128;
  Best Local Similarity  66.7%;  Pred. No. 2.31e+01;
  Matches    6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db     15 YLATASTMD 23
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT   15
ENTRY           A37246     #type complete
TITLE           myelin basic protein - guinea pig
ALTERNATE_NAMES myelin A1 protein
ORGANISM        #formal_name Cavia porcellus #common_name guinea pig
DATE            31-Jul-1991 #sequence_revision 31-Jul-1991 #text_change
```

```
                  07-Oct-1994
ACCESSIONS        A37246; C92087; A03140
REFERENCE         A37246
   #authors       Deibler, G.E.; Martenson, R.E.; Krutzsch, H.C.; Kies, M.W.
   #journal       J. Neurochem. (1984) 43:100-105
   #title         Sequence of guinea pig myelin basic protein.
   #cross-references MUID:84215086
   #accession     A37246
      ##status         preliminary
      ##molecule_type protein
      ##residues       1-167 ##label DEI
REFERENCE         A92087
   #authors       Shapira, R.; McKneally, S.S.; Chou, F.; Kibler, R.F.
   #journal       J. Biol. Chem. (1971) 246:4630-4640
   #title         Encephalitogenic fragment of myelin basic protein. Amino acid
                   sequence of bovine, rabbit, guinea pig, monkey, and human
                   fragments.
   #accession     C92087
      ##molecule_type protein
      ##residues       45-87 ##label SHA
CLASSIFICATION    #superfamily myelin basic protein
KEYWORDS          myelin
SUMMARY           #length 167  #molecular-weight 18213  #checksum 1628

  Query Match           71.2%;  Score 47;  DB 2;  Length 167;
  Best Local Similarity  66.7%;  Pred. No. 2.31e+01;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db      14 YLATASTMD 22
           ||:|||::|
Qy       3 YLSTASSLD 11


Search completed: Thu Sep  2 11:21:08 1999
Job time : 14 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 11:22:32 1999; MasPar time 1.38 Seconds
88.032 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-4
Description: (1-12) from US08599226.pep
Perfect Score: 66
Sequence: 1 VSYLSTASSLDX 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 15.730; Variance 46.168; scale 0.341

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	48	72.7	426	3	PCT-US95-1	Sequence 2, Applicatio	5.13e+01
2	48	72.7	426	1	US-08-336-	Sequence 2, Applicatio	5.13e+01
3	47	71.2	20	2	US-08-468-	Sequence 2, Applicatio	6.62e+01
4	47	71.2	21	2	US-08-787-	Sequence 33, Applicati	6.62e+01
5	47	71.2	168	4	5194425-4	Patent No. 5194425.	6.62e+01
6	47	71.2	170	2	US-08-327-	Sequence 1, Applicatio	6.62e+01
7	45	68.2	345	2	US-08-446-	Sequence 40, Applicati	1.10e+02
8	45	68.2	384	1	US-08-457-	Sequence 5, Applicatio	1.10e+02
9	44	66.7	170	4	5194425-3	Patent No. 5194425.	1.41e+02
10	44	66.7	170	2	US-08-227-	Sequence 1, Applicatio	1.41e+02
11	43	65.2	209	2	US-08-771-	Sequence 2, Applicatio	1.81e+02
12	43	65.2	822	2	US-08-222-	Sequence 7, Applicatio	1.81e+02
13	43	65.2	913	1	US-08-220-	Sequence 6, Applicatio	1.81e+02
14	43	65.2	913	1	US-08-413-	Sequence 6, Applicatio	1.81e+02
15	43	65.2	1255	2	US-08-356-	Sequence 2, Applicatio	1.81e+02
16	43	65.2	1255	2	US-08-484-	Sequence 8, Applicatio	1.81e+02
17	43	65.2	1255	2	US-08-414-	Sequence 68, Applicati	1.81e+02
18	43	65.2	1255	2	US-08-486-	Sequence 68, Applicati	1.81e+02
19	43	65.2	1255	2	US-08-625-	Sequence 2, Applicatio	1.81e+02
20	43	65.2	1255	2	US-08-468-	Sequence 68, Applicati	1.81e+02
21	43	65.2	1255	1	US-08-467-	Sequence 68, Applicati	1.81e+02
22	43	65.2	3666	2	US-08-222-	Sequence 12, Applicati	1.81e+02
23	43	65.2	3727	2	US-08-222-	Sequence 27, Applicati	1.81e+02
24	43	65.2	3778	2	US-08-222-	Sequence 2, Applicatio	1.81e+02
25	42	63.6	31	3	PCT-US93-0	Sequence 111, Applicat	2.31e+02
26	42	63.6	31	1	US-08-118-	Sequence 111, Applicat	2.31e+02
27	42	63.6	43	3	PCT-US93-0	Sequence 248, Applicat	2.31e+02
28	42	63.6	43	1	US-08-118-	Sequence 248, Applicat	2.31e+02
29	42	63.6	166	3	PCT-US96-0	Sequence 9, Applicatio	2.31e+02
30	42	63.6	166	3	PCT-US96-0	Sequence 2, Applicatio	2.31e+02
31	42	63.6	170	4	5468481-3	Patent No. 5468481.	2.31e+02
32	42	63.6	238	3	PCT-US95-1	Sequence 2, Applicatio	2.31e+02
33	42	63.6	238	2	US-08-753-	Sequence 2, Applicatio	2.31e+02
34	42	63.6	238	1	US-08-452-	Sequence 1, Applicatio	2.31e+02
35	42	63.6	238	2	US-08-588-	Sequence 2, Applicatio	2.31e+02
36	42	63.6	238	1	US-08-337-	Sequence 2, Applicatio	2.31e+02
37	42	63.6	326	4	5395759-2	Patent No. 5395759.	2.31e+02
38	42	63.6	335	2	US-09-057-	Sequence 7, Applicatio	2.31e+02
39	42	63.6	403	2	US-08-712-	Sequence 3, Applicatio	2.31e+02
40	42	63.6	426	2	US-08-852-	Sequence 2, Applicatio	2.31e+02
41	42	63.6	464	1	US-07-688-	Sequence 16, Applicati	2.31e+02
42	42	63.6	464	3	PCT-US91-0	Sequence 16, Applicati	2.31e+02
43	42	63.6	980	1	US-08-220-	Sequence 5, Applicatio	2.31e+02
44	42	63.6	980	1	US-08-413-	Sequence 5, Applicatio	2.31e+02
45	41	62.1	192	3	PCT-US95-1	Sequence 101, Applicat	2.95e+02

ALIGNMENTS

```
RESULT    1
ID   PCT-US95-13795-2     STANDARD;     PRT;   426 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9513795
XX
CC    Sequence 2, Application PC/TUS9513795
CC     GENERAL INFORMATION:
CC       APPLICANT:  HOLLIS, GREGORY F.
CC       APPLICANT:  PATEL, MAYUR D.
CC       TITLE OF INVENTION:  DNA ENCODING CANINE IMMUNOGLOBULINS
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  126 E. LINCOLN AVENUE; P.O. BOX 2000
CC         CITY:  RAHWAY
CC         STATE:  NEW JERSEY
CC         COUNTRY:  USA
CC         ZIP:  07065-0907
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/13795
CC         FILING DATE:
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  19211Y
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (908) 594-6734
CC         TELEFAX:  (908) 594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  426 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   426 AA; 47234 MW; 1032622 CN;
```

```
  Query Match           72.7%;  Score 48;  DB 3;  Length 426;
  Best Local Similarity  70.0%;  Pred. No. 5.13e+01;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    215 SYLSPPSPLD 224
          |||| :|:||
Qy      2 SYLSTASSLD 11


RESULT    2
ID   US-08-336-583-2     STANDARD;     PRT;   426 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08336583
XX
CC    Sequence 2, Application US/08336583
CC    Patent No. 5629415
CC     GENERAL INFORMATION:
CC       APPLICANT:  HOLLIS, GREGORY F.
CC       APPLICANT:  PATEL, MAYUR D.
CC       TITLE OF INVENTION:  DNA ENCODING CANINE IMMUNOGLOBULIN E
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  126 E. LINCOLN AVENUE
CC         CITY:  RAHWAY
CC         STATE:  NEW JERSEY
CC         COUNTRY:  USA
CC         ZIP:  07065-0900
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/336,583
CC         FILING DATE:  09-NOV-1994
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  19211
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (908) 594-6734
CC         TELEFAX:  (908) 594-4720
CC      INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  426 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   426 AA; 47234 MW; 1032622 CN;

  Query Match           72.7%;  Score 48;  DB 1;  Length 426;
  Best Local Similarity  70.0%;  Pred. No. 5.13e+01;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    215 SYLSPPSPLD 224
          |||| :|:||
Qy      2 SYLSTASSLD 11


RESULT    3
ID   US-08-468-540B-2     STANDARD;     PRT;   20 AA.
XX
AC   xxxxxx
XX
DT
XX
```

```
DE   Sequence 2, Application US/08468540B
XX
CC    Sequence 2, Application US/08468540B
CC    Patent No. 5858980
CC     GENERAL INFORMATION:
CC       APPLICANT:  Weiner, Howard
CC       APPLICANT:  Hafler, David
CC       APPLICANT:  Miller, Ariel
CC       APPLICANT:  Al-Sabbagh, Ahmad
CC       TITLE OF INVENTION:  SUPPRESSION OF T-CELL PROLIFERATION
CC       TITLE OF INVENTION:  USING PEPTIDE FRAGMENTS OF MYELIN BASIC PROTEIN
CC       NUMBER OF SEQUENCES:  25
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Darby & Darby P.C.
CC         STREET:  805 Third Avenue
CC         CITY:  New York
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10022
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/468,540B
CC         FILING DATE:  06-JUN-1995
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Jacobs, Seth H
CC         REGISTRATION NUMBER:  32,140
CC         REFERENCE/DOCKET NUMBER:
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-527-7700
CC         TELEFAX:
CC         TELEX:
CC      INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  20 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  No. 5858980e
SQ   SEQUENCE   20 AA; 2160 MW; 2047 CN;

  Query Match           71.2%;  Score 47;  DB 2;  Length 20;
  Best Local Similarity  66.7%;  Pred. No. 6.62e+01;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db      4 YLATASTMD 12
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT    4
ID   US-08-787-547-33     STANDARD;     PRT;   21 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 33, Application US/08787547
XX
CC    Sequence 33, Application US/08787547
CC    Patent No. 5783567
CC     GENERAL INFORMATION:
CC       APPLICANT:  Hedley, Mary Lynne
CC       APPLICANT:  Curley, Joanne M.
CC       APPLICANT:  Langer, Robert S.
CC       TITLE OF INVENTION:  MICROPARTICLES FOR DELIVERY
```

```
CC       TITLE OF INVENTION:  OF NUCLEIC ACID
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson, P.C.
CC         STREET:  225 Franklin Street
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  US
CC         ZIP:  02110-2804
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  Windows95
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/787,547
CC         FILING DATE:  22-JAN-1997
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Fraser, Janis K.
CC         REGISTRATION NUMBER:  34,819
CC         REFERENCE/DOCKET NUMBER:  08191/003001
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  617-542-5070
CC         TELEFAX:  617-542-8906
CC         TELEX:  200154
CC     INFORMATION FOR SEQ ID NO:  33:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  21 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   21 AA; 2312 MW; 2012 CN;

  Query Match           71.2%;  Score 47;  DB 2;  Length 21;
  Best Local Similarity  66.7%;  Pred. No. 6.62e+01;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db      8 YLATASTMD 16
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT    5
ID   5194425-4     STANDARD;     PRT;   182 AA.
XX
AC   xxxxxx
XX
DT   01-JAN-1900
XX
DE   Patent No. 5194425.
XX
CC   Patent No. 5194425
CC       APPLICANT: SHARMA ,SOMESH D.;LERCH, L. BERNARD;CLARK,
CC   BRIAN R.
CC       TITLE OF INVENTION: MHC-MEDIATED TOXIC CONJUGATES USEFUL IN
CC   AMELIORATING AUTOIMMUNITY
CC       NUMBER OF SEQUENCES: 9
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/367,751
CC         FILING DATE: 21-JUN-1989
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER: 210,594
CC         FILING DATE: 23-JUN-1988
CC   SEQ ID NO:4:
CC         LENGTH: 168
SQ   SEQUENCE   182 AA; 19707 MW; 182144 CN;

  Query Match           71.2%;  Score 47;  DB 4;  Length 168;
  Best Local Similarity  66.7%;  Pred. No. 6.62e+01;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db     12 YLATASTMD 20
          ||:|||::|
Qy      3 YLSTASSLD 11


RESULT    6
ID   US-08-327-357A-1     STANDARD;     PRT;   170 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08327357A
XX
CC    Sequence 1, Application US/08327357A
CC    Patent No. 5817629
CC     GENERAL INFORMATION:
CC       APPLICANT:  WARREN, Kenneth G.
CC       APPLICANT:  CATZ, Ingrid
CC       TITLE OF INVENTION:  PEPTIDE SPECIFICITY OF ANTI-MYELIN BASIC
CC       TITLE OF INVENTION:  PROTEIN AND THE ADMINISTRATION OF MYELIN BASIC PROTEIN
CC       TITLE OF INVENTION:  PEPTIDES TO MUTIPLE SCLEROSIS PATIENTS
CC       NUMBER OF SEQUENCES:  1
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Venable, Baetjer, Howard & Civiletti
CC         STREET:  1201 New York Avenue, N.W., Suite 1000
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20005
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/327,357A
CC         FILING DATE:  21-OCT-1994
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/798,099
CC         FILING DATE:  27-NOV-1991
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  CA 2,053,799-0
CC         FILING DATE:  22-OCT-1991
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Ihnen, Jeffrey L.
CC         REGISTRATION NUMBER:  28,957
CC         REFERENCE/DOCKET NUMBER:  27052-115469
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-962-4810
CC         TELEFAX:  202-962-8300
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  170 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Homo sapiens
CC       IMMEDIATE SOURCE:
CC         CLONE:  human myelin basic protein
SQ   SEQUENCE   170 AA; 18459 MW; 143992 CN;

  Query Match           71.2%;  Score 47;  DB 2;  Length 170;
  Best Local Similarity  66.7%;  Pred. No. 6.62e+01;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db     14 YLATASTMD 22
          ||:|||::|
```

```
Qy      3 YLSTASSLD 11


RESULT    7
ID   US-08-446-345-40     STANDARD;     PRT;   345 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 40, Application US/08446345
XX
CC    Sequence 40, Application US/08446345
CC    Patent No. 5831009
CC     GENERAL INFORMATION:
CC       APPLICANT:  Ullrich, Axel
CC       APPLICANT:  Moller, Niels P.H.
CC       APPLICANT:  Moller, Karin B.
CC       TITLE OF INVENTION:  NOVEL PROTEIN PHOSPHOTYROSINE
CC       TITLE OF INVENTION:  PHOSPHATASES PTP-D1
CC       NUMBER OF SEQUENCES:  41
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  N.Y.
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/446,345
CC         FILING DATE:  22-MAY-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/234,440
CC         FILING DATE:  28-APR-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Coruzzi, Laura A.
CC         REGISTRATION NUMBER:  30742
CC         REFERENCE/DOCKET NUMBER:  7683-054
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212)790-9090
CC         TELEFAX:  (212) 869-8864
CC         TELEX:  66141 PENNIE
CC      INFORMATION FOR SEQ ID NO:  40:
CC        SEQUENCE CHARACTERISTICS:
CC         LENGTH:  345 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC        MOLECULE TYPE:  protein
SQ   SEQUENCE   345 AA; 40427 MW; 614554 CN;

  Query Match           68.2%;  Score 45;  DB 2;  Length 345;
  Best Local Similarity  50.0%;  Pred. No. 1.10e+02;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    205 NYLNTARTLE 214
          :||:|| :|:
Qy      2 SYLSTASSLD 11


RESULT    8
ID   US-08-457-245-5     STANDARD;     PRT;   384 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/08457245
XX
CC    Sequence 5, Application US/08457245
CC    Patent No. 5573915
CC     GENERAL INFORMATION:
CC       APPLICANT:  BARRY III, Clifton E.
CC       APPLICANT:  YUAN, Ying
CC       TITLE OF INVENTION:  CLONING AND EXPRESSION OF DNA INVOLVED
CC       TITLE OF INVENTION:  IN THE BIOSYNTHESIS OF CYCLOPROPANATED MYCOLIC ACIDS IN
CC       TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS
CC       NUMBER OF SEQUENCES:  21
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC         STREET:  Steuart Street Tower, One Market Plaza
CC         CITY:  San Francisco
CC         STATE:  California
CC         COUNTRY:  US
CC         ZIP:  94105-1493
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/457,245
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Chambers, Guy W.
CC         REGISTRATION NUMBER:  30,617
CC         REFERENCE/DOCKET NUMBER:  15280-216000
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 543-9600
CC         TELEFAX:  (415) 543-5043
CC      INFORMATION FOR SEQ ID NO:  5:
CC        SEQUENCE CHARACTERISTICS:
CC         LENGTH:  384 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC        MOLECULE TYPE:  ORF2 protein
SQ   SEQUENCE   384 AA; 41963 MW; 701271 CN;

  Query Match           68.2%;  Score 45;  DB 1;  Length 384;
  Best Local Similarity  60.0%;  Pred. No. 1.10e+02;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    236 SYLPTKAALD 245
          |||:| ::||
Qy      2 SYLSTASSLD 11


RESULT    9
ID   5194425-3     STANDARD;     PRT;   184 AA.
XX
AC   xxxxxx
XX
DT   01-JAN-1900
XX
DE   Patent No. 5194425.
XX
CC   Patent No. 5194425
CC       APPLICANT: SHARMA ,SOMESH D.;LERCH, L. BERNARD;CLARK,
CC   BRIAN R.
CC       TITLE OF INVENTION: MHC-MEDIATED TOXIC CONJUGATES USEFUL IN
CC   AMELIORATING AUTOIMMUNITY
CC       NUMBER OF SEQUENCES: 9
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/367,751
CC         FILING DATE: 21-JUN-1989
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER: 210,594
CC         FILING DATE: 23-JUN-1988
CC   SEQ ID NO:3:
```

```
CC         LENGTH: 170
SQ   SEQUENCE   184 AA; 19952 MW; 183015 CN;

  Query Match           66.7%;  Score 44;  DB 4;  Length 170;
  Best Local Similarity  55.6%;  Pred. No. 1.41e+02;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     13 YLASASTMD 21
          ||::||::|
Qy      3 YLSTASSLD 11


RESULT   10
ID   US-08-227-372-1     STANDARD;     PRT;   170 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08227372

     Sequence 1, Application US/08227372
CC   Patent No. 5763585
CC    GENERAL INFORMATION:
CC      APPLICANT:  Nag, Bishwajit
CC      TITLE OF INVENTION:  PURIFICATION AND CHARACTERIZATION OF
CC      TITLE OF INVENTION:  MHC-PEPTIDE COMPLEXES USEFUL IN AMELIORATING AUTOIMMUNITY
CC      NUMBER OF SEQUENCES:  4
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/227,372
CC        FILING DATE:  14-APR-1994
CC        CLASSIFICATION:  424
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  US 08/136,216
CC        FILING DATE:  13-OCT-1993
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  14058-32-1
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC    INFORMATION FOR SEQ ID NO:  1:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  170 amino acids
CC        TYPE:  amino acid
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  protein
CC      FEATURE:
CC        NAME/KEY:  Protein
CC        LOCATION:  1..170
CC        OTHER INFORMATION:  /note= "Myelin basic protein"
SQ   SEQUENCE   170 AA; 18410 MW; 141752 CN;

  Query Match           66.7%;  Score 44;  DB 2;  Length 170;
  Best Local Similarity  55.6%;  Pred. No. 1.41e+02;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     13 YLASASTMD 21
          ||::||::|
Qy      3 YLSTASSLD 11


RESULT   11
ID   US-08-771-784-2     STANDARD;     PRT;   209 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08771784
XX
CC    Sequence 2, Application US/08771784
CC    Patent No. 5846772
CC     GENERAL INFORMATION:
CC       APPLICANT:  John E. Hodgson
CC       APPLICANT:  Nicola G. Wallis
CC       TITLE OF INVENTION:  NOVEL TCTS RESPONSE REGUL
CC       TITLE OF INVENTION:  ATOR
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  SmithKline Beecham Corporation
CC         STREET:  709 Swedeland Road
CC         CITY:  King of Prussia
CC         STATE:  PA
CC         COUNTRY:  USA
CC         ZIP:  19406-0939
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/771,784
CC         FILING DATE:  20-DEC-1996
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  9526352.1
CC         FILING DATE:  22-DEC-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Gimmi, Edward R
CC         REGISTRATION NUMBER:  38,891
CC         REFERENCE/DOCKET NUMBER:  P31323
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  610-270-4478
CC         TELEFAX:  610-270-5090
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  209 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   209 AA; 23559 MW; 225335 CN;

  Query Match           65.2%;  Score 43;  DB 2;  Length 209;
  Best Local Similarity  60.0%;  Pred. No. 1.81e+02;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db     20 SYLSTQSDIE 29
          ||||| | ::
Qy      2 SYLSTASSLD 11


RESULT   12
ID   US-08-222-617A-7     STANDARD;     PRT;   822 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 7, Application US/08222617A
XX
```

```
CC    Sequence 7, Application US/08222617A
CC    Patent No. 5882879
CC     GENERAL INFORMATION:
CC       APPLICANT:  Veenstra, Annemarie E.
CC       APPLICANT:  Martin, Juan F.
CC       APPLICANT:  Garcia, Bruno D.
CC       APPLICANT:  Gutierrez, Santiago
CC       APPLICANT:  Barredo, Jose L.
CC       APPLICANT:  Von Doehren, Hans
CC       APPLICANT:  Palissa, Harriet
CC       APPLICANT:  Van Liempt, Henk
CC       APPLICANT:  Montenegro, Eduardo P.
CC       TITLE OF INVENTION:  A Method for Influencing Beta-Lactam
CC       TITLE OF INVENTION:  Antibiotic Production and for Isolation of Large
CC       TITLE OF INVENTION:  Quantities of ACV Synthetase
CC       NUMBER OF SEQUENCES:  27
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
CC         STREET:  300 South Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/222,617A
CC         FILING DATE:  04-APR-1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         REFERENCE/DOCKET NUMBER:  97,157
CC     INFORMATION FOR SEQ ID NO:  7:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  822 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       FRAGMENT TYPE:  internal
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Penicillium chrysogenum
CC       FEATURE:
CC         NAME/KEY:  Domain
CC         LOCATION:  1..822
CC         OTHER INFORMATION:  /label= Domain III
CC         OTHER INFORMATION:  /note= "Domain III of ACV Synthetase from
CC         OTHER INFORMATION:  Penicillium chrysogenum;  aa 2474-3295"
SQ   SEQUENCE   822 AA; 91699 MW; 3418282 CN;

  Query Match           65.2%;  Score 43;  DB 2;  Length 822;
  Best Local Similarity  60.0%;  Pred. No. 1.81e+02;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    267 LSYLSGTPSL 276
          :|||| ::||
Qy      1 VSYLSTASSL 10


RESULT   13
ID   US-08-220-151-6     STANDARD;     PRT;   913 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08220151
XX
CC    Sequence 6, Application US/08220151
CC    Patent No. 5529780
CC     GENERAL INFORMATION:
CC       APPLICANT:  Paoletti, Enzo
CC       APPLICANT:  Limbach, Keith J.
CC       TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES OF
CC       TITLE OF INVENTION:  CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR
CC       NUMBER OF SEQUENCES:  91
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Curtis, Morris & Safford
CC         STREET:  530 Fifth Avenue
CC         CITY:  New York
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/220,151
CC         FILING DATE:  30-MAR-1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Frommer, William S.
CC         REGISTRATION NUMBER:  25,506
CC         REFERENCE/DOCKET NUMBER:  454310-2540
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 840-3333
CC         TELEFAX:  (212) 840-0712
CC         TELEX:  425066 CURTMS
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  913 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       FRAGMENT TYPE:  N-terminal
SQ   SEQUENCE   913 AA; 100233 MW; 4073940 CN;

  Query Match           65.2%;  Score 43;  DB 1;  Length 913;
  Best Local Similarity  36.4%;  Pred. No. 1.81e+02;
  Matches     4;  Conservative    4;  Mismatches   3;  Indels   0;  Gaps   0;

Db    862 IRYMSIVSALE 872
          : |:|  |:|:
Qy      1 VSYLSTASSLD 11


RESULT   14
ID   US-08-413-118-6     STANDARD;     PRT;   913 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08413118
XX
CC    Sequence 6, Application US/08413118
CC    Patent No. 5688920
CC     GENERAL INFORMATION:
CC       APPLICANT:  PAOLETTI, ENZO
CC       APPLICANT:  LIMBACH, KEITH J.
CC       TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES OF
CC       TITLE OF INVENTION:  CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR
CC       NUMBER OF SEQUENCES:  128
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CURTIS, MORRIS & SAFFORD, P.C.
CC         STREET:  530 FIFTH AVENUE, 25TH FLOOR
CC         CITY:  NEW YORK
CC         STATE:  NEW YORK
CC         COUNTRY:  UNITED STATES OF AMERICA
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
```

```
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/413,118
CC         FILING DATE:  29-MAR-1995
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/220,151
CC         FILING DATE:  30-MAR-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  FROMMER, WILLIAM S.
CC         REGISTRATION NUMBER:  25,506
CC         REFERENCE/DOCKET NUMBER:  454310-2670
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 840-3333
CC         TELEFAX:  (212) 840-0712
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  913 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       FRAGMENT TYPE:  N-terminal
SQ   SEQUENCE   913 AA; 100233 MW; 4073940 CN;

  Query Match           65.2%;  Score 43;  DB 1;  Length 913;
  Best Local Similarity  36.4%;  Pred. No. 1.81e+02;
  Matches     4;  Conservative    4;  Mismatches   3;  Indels   0;  Gaps   0;

Db    862 IRYMSIVSALE 872
          : |:|  |:|:
Qy      1 VSYLSTASSLD 11


RESULT   15
ID   US-08-356-786-2     STANDARD;     PRT;   1255 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08356786
XX
CC    Sequence 2, Application US/08356786
CC    Patent No. 5877305
CC     GENERAL INFORMATION:
CC       APPLICANT:  Huston, James S.
CC       APPLICANT:  Oppermann, Hermann
CC       APPLICANT:  Houston, L. L.
CC       APPLICANT:  Ring, David B.
CC       TITLE OF INVENTION:  Biosynthetic Binding Protein for Cancer
CC       TITLE OF INVENTION:  Marker
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Edmund R. Pitcher, Testa, Hurwitz, & Thibeault
CC         STREET:  Exchange Place, 53 State Street
CC         CITY:  Boston
CC         STATE:  Massachusetts
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/356,786
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/831,967
CC         FILING DATE:  06-FEB-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Pitcher, Edmund R.
CC         REGISTRATION NUMBER:  27,829
CC         REFERENCE/DOCKET NUMBER:  CRP-053
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617) 248-7000
CC         TELEFAX:  (617) 248-7100
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1255 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1255 AA; 137909 MW; 8111405 CN;

  Query Match           65.2%;  Score 43;  DB 2;  Length 1255;
  Best Local Similarity  50.0%;  Pred. No. 1.81e+02;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db     62 LTYLPTNASL 71
          ::||:| :||
Qy      1 VSYLSTASSL 10


Search completed: Thu Sep  2 11:22:39 1999
Job time : 7 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 11:18:01 1999; MasPar time 2.14 Seconds
118.732 Million cell updates/sec

Tabular output not generated.

Title: >US-08-599-226-3
Description: (1-9) from US08599226.pep
Perfect Score: 66
Sequence: 1 QRYNRAPYX 9

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 23.600; Variance 27.313; scale 0.864

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	57	86.4	279	1	PRC2_DROME	PROTEASOME 29 KD SUBUN	8.67e-02
2	54	81.8	263	1	PRC2_RAT	PROTEASOME COMPONENT C	4.20e-01
3	54	81.8	263	1	PRC2_HUMAN	PROTEASOME COMPONENT C	4.20e-01
4	51	77.3	1262	1	GNRP_MOUSE	GUANINE NUCLEOTIDE REL	1.92e+00
NR> 5	50	75.8	287	1	FCAR_HUMAN	IMMUNOGLOBULIN ALPHA F	3.15e+00
6	49	74.2	360	1	DCAM_SOLTU	S-ADENOSYLMETHIONINE D	5.13e+00
7	48	72.7	285	1	RS2_ARATH	40S RIBOSOMAL PROTEIN	8.28e+00
8	47	71.2	14	1	HY14_PIG	HYPOTHALAMIC TETRADECA	1.33e+01
9	47	71.2	288	1	TRA6_NEIME	TRANSPOSASE FOR INSERT	1.33e+01
10	47	71.2	323	1	GDIA_MOUSE	RAB GDP DISSOCIATION I	1.33e+01
11	47	71.2	442	1	VIT2_DROME	VITELLOGENIN II PRECUR	1.33e+01
12.	47	71.2	445	1	GDIB_HUMAN	RAB GDP DISSOCIATION I	1.33e+01
13	47	71.2	445	1	GDIB_RAT	RAB GDP DISSOCIATION I	1.33e+01
14	47	71.2	445	1	GDIB_MOUSE	RAB GDP DISSOCIATION I	1.33e+01
15	47	71.2	447	1	GDIA_HUMAN	RAB GDP DISSOCIATION I	1.33e+01
16	47	71.2	447	1	GDIA_BOVIN	RAB GDP DISSOCIATION I	1.33e+01
17	47	71.2	447	1	GDIA_RAT	RAB GDP DISSOCIATION I	1.33e+01
18	47	71.2	451	1	GDI1_YEAST	SECRETORY PATHWAY GDP	1.33e+01
19	47	71.2	689	1	YBP3_YEAST	HYPOTHETICAL 77.3 KD P	1.33e+01
20	47	71.2	1320	1	PUTA_ECOLI	PROLINE DEHYDROGENASE	1.33e+01
21	47	71.2	1320	1	PUTA_SALTY	PROLINE DEHYDROGENASE	1.33e+01
22	47	71.2	3305	1	APLP_MANSE	APOLIPOPHORIN PRECURSO	1.33e+01
23	46	69.7	182	1	HPRT_ECOLI	HYPOXANTHINE PHOSPHORI	2.11e+01
24	46	69.7	360	1	FARX_LYMST	FMRFAMIDE-RELATED NEUR	2.11e+01
25	46	69.7	365	1	NADA_SALTY	QUINOLINATE SYNTHETASE	2.11e+01
26	46	69.7	610	1	EDD_PSEAE	PHOSPHOGLUCONATE DEHYD	2.11e+01
27	46	69.7	1385	1	YMS5_CAEEL	HYPOTHETICAL 159.2 KD	2.11e+01
28	45	68.2	108	1	KV1L_HUMAN	IG KAPPA CHAIN V-I REG	3.33e+01
29	45	68.2	255	1	YAFJ_ECOLI	HYPOTHETICAL 28.6 KD P	3.33e+01
30	45	68.2	406	1	BHMT_HUMAN	BETAINE--HOMOCYSTEINE	3.33e+01
31	45	68.2	407	1	BHMT_RAT	BETAINE--HOMOCYSTEINE	3.33e+01
32	45	68.2	423	1	TRPB_HALVO	TRYPTOPHAN SYNTHASE BE	3.33e+01
33	45	68.2	467	1	MATK_RAT	MEGAKARYOCYTE-ASSOCIAT	3.33e+01
34	45	68.2	491	1	CG2A_DROME	G2/MITOTIC-SPECIFIC CY	3.33e+01
35	45	68.2	494	1	PYR5_NAEGR	URIDINE 5'-MONOPHOSPHA	3.33e+01
36	45	68.2	505	1	MATK_MOUSE	MEGAKARYOCYTE-ASSOCIAT	3.33e+01
37	45	68.2	507	1	MATK_HUMAN	MEGAKARYOCYTE-ASSOCIAT	3.33e+01
38	45	68.2	521	1	NPRE_BACAM	BACILLOLYSIN PRECURSOR	3.33e+01
39	45	68.2	563	1	LIP2_GEOCN	LIPASE 2 PRECURSOR (EC	3.33e+01
40	45	68.2	563	1	LIP1_GEOCN	LIPASE 1 PRECURSOR (EC	3.33e+01
41	45	68.2	616	1	HSCA_ECOLI	CHAPERONE PROTEIN HSCA	3.33e+01
42	45	68.2	2410	1	POL1_BAYMJ	GENOME POLYPROTEIN 1 [	3.33e+01
43	45	68.2	2412	1	POL1_BAYMG	GENOME POLYPROTEIN 1 [	3.33e+01
44	44	66.7	245	1	Y382_METJA	HYPOTHETICAL PROTEIN M	5.22e+01
45	44	66.7	615	1	ALBU_CHICK	SERUM ALBUMIN PRECURSO	5.22e+01

ALIGNMENTS

RESULT 1
ID PRC2_DROME STANDARD; PRT; 279 AA.
AC P12881;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROTEASOME 29 KD SUBUNIT (EC 3.4.99.46) (MULTICATALYTIC ENDOPEPTIDASE
DE COMPLEX 35 KD SUBUNIT).
GN PROS35 OR PROS-35.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 4-18 AND 194-206.
RC STRAIN=OREGON-R;
RX MEDLINE; 90005444.
RA HAASS C., PESOLD-HURT B., MULTHAUP G., BEYREUTHER K.,
RA KLOETZEL P.-M.;
RT "The PROS-35 gene encodes the 35 kd protein subunit of Drosophila
RT melanogaster proteasome.";
RL EMBO J. 8:2373-2379(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE; 92249308.
RA FRENTZEL S., TROXELL M., HAASS C., PESOLD-HURT B., GLAETZER K.H.,
RA KLOETZEL P.-M.;
RT "Molecular characterization of the genomic regions of the Drosophila
RT alpha-type subunit proteasome genes PROS-Dm28.1 and PROS-Dm35.";
RL EUR. J. BIOCHEM. 205:1043-1051(1992).
CC -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
CC WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH
CC ARG, PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT
CC NEUTRAL OR SLIGHTLY BASIC PH. THE 35 KD SUBUNIT IS PROBABLY A
CC REGULATORY SUBUNIT. THE PROTEASOME HAS AN ATP-DEPENDENT
CC PROTEOLYTIC ACTIVITY.
CC -!- PATHWAY: IS INVOLVED IN AN ATP/UBIQUITIN-DEPENDENT NON-LYSOSOMAL
CC PROTEOLYTIC PATHWAY.
CC -!- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
CC SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE.
CC -!- SUBCELLULAR LOCATION: PROTEASOMES ARE FOUND IN THE CYTOPLASM AND
CC ALSO IN THE NUCLEUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A; ALSO KNOWN AS THE
CC PROTEASOME A-TYPE FAMILY. BELONGS TO THE C2 SUBFAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X15497; G8382; -.
DR   EMBL; X62285; G8388; -.
DR   PIR; S05507; SNFF5K.
DR   PIR; S23450; S23450.
DR   FLYBASE; FBgn0003151; Pros35.
DR   PROSITE; PS00388; PROTEASOME_A; 1.
DR   PFAM; PF00227; proteasome; 1.
DR   HSSP; P25156; 1PMA.
KW   PROTEASOME; HYDROLASE; PROTEASE; PHOSPHORYLATION.
FT   MOD_RES     103    103       PHOSPHORYLATION (POTENTIAL).
SQ   SEQUENCE   279 AA;  31058 MW;  AA18BBDF CRC32;

  Query Match           86.4%;  Score 57;  DB 1;  Length 279;
  Best Local Similarity  75.0%;  Pred. No. 8.67e-02;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    121 QRYDRRPY 128
          |||:| ||
Qy      1 QRYNRAPY 8


RESULT    2
ID   PRC2_RAT       STANDARD;      PRT;   263 AA.
AC   P18420;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PROTEASOME COMPONENT C2 (EC 3.4.99.46) (MACROPAIN SUBUNIT C2)
DE   (PROTEASOME NU CHAIN) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT
DE   C2).
GN   PSMA1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 90057428.
RA   FUJIWARA T., TANAKA K., KUMATORI A., SHIN S., YOSHIMURA T.,
RA   ICHIHARA A., TOKUNAGA F., ARUGA R., IWANAGA S., KAKIZUKA A.,
RA   NAKANISHI S.;
RT   "Molecular cloning of cDNA for proteasomes (multicatalytic proteinase
RT   complexes) from rat liver: primary structure of the largest component
RT   (C2).";
RL   BIOCHEMISTRY 28:7332-7340(1989).
RN   [2]
RP   SEQUENCE OF 1-30.
RC   TISSUE=LIVER;
RX   MEDLINE; 90243011.
RA   TOKUNAGA F., ARUGA R., IWANAGA S., TANAKA K., ICHIHARA A., TAKAO T.,
RA   SHIMONISHI Y.;
RT   "The NH2-terminal residues of rat liver proteasome (multicatalytic
RT   proteinase complex) subunits, C2, C3 and C8, are N
RT   alpha-acetylated.";
RL   FEBS LETT. 263:373-375(1990).
CC   -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
CC       WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH
CC       ARG, PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT
CC       NEUTRAL OR SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT
CC       PROTEOLYTIC ACTIVITY.
CC   -!- PATHWAY: IS INVOLVED IN AN ATP/UBIQUITIN-DEPENDENT NON-LYSOSOMAL
CC       PROTEOLYTIC PATHWAY.
CC   -!- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
CC       SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: PROTEASOMES ARE FOUND IN THE CYTOPLASM AND
CC       ALSO IN THE NUCLEUS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ALL RAT TISSUES.
CC   -!- PTM: ITS C-TERMINAL EXTENSION IS PARTIALLY CLEAVED OFF BY LIMITED
CC       PROTEOLYSIS LEADING TO A CONVERSION OF THE PROTEASOME FROM ITS
CC       LATENT INTO ITS ACTIVE FORM.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A; ALSO KNOWN AS THE
CC       PROTEASOME A-TYPE FAMILY. BELONGS TO THE C2 SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M29859; G206382; -.
DR   EMBL; D90265; G220877; -.
DR   PIR; A32968; SNRTC2.
DR   PROSITE; PS00388; PROTEASOME_A; 1.
DR   PFAM; PF00227; proteasome; 1.
DR   HSSP; P25156; 1PMA.
KW   PROTEASOME; HYDROLASE; PROTEASE; ACETYLATION.
FT   MOD_RES       1      1       ACETYLATION.
SQ   SEQUENCE   263 AA;  29517 MW;  01D55620 CRC32;

  Query Match           81.8%;  Score 54;  DB 1;  Length 263;
  Best Local Similarity  75.0%;  Pred. No. 4.20e-01;
  Matches     6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    121 QRYGRRPY 128
          ||| | ||
Qy      1 QRYNRAPY 8


RESULT    3
ID   PRC2_HUMAN     STANDARD;      PRT;   263 AA.
AC   P25786;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PROTEASOME COMPONENT C2 (EC 3.4.99.46) (MACROPAIN SUBUNIT C2)
DE   (PROTEASOME NU CHAIN) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT
DE   C2) (30 KD PROSOMAL PROTEIN) (PROS-30).
GN   PSMA1 OR PSC2 OR PROS30.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93013039.
RA   SILVA-PEREIRA I., BEY F., COUX O., SCHERRER K.;
RT   "Two mRNAs exist for the Hs PROS-30 gene encoding a component of
RT   human prosomes.";
RL   GENE 120:235-242(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91223105.
RA   TAMURA T., LEE D.H., OSAKA F., FUJIWARA T., SHIN S., CHUNG C.H.,
RA   TANAKA K., ICHIHARA A.;
RT   "Molecular cloning and sequence analysis of cDNAs for five major
RT   subunits of human proteasomes (multi-catalytic proteinase
RT   complexes).";
RL   BIOCHIM. BIOPHYS. ACTA 1089:95-102(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91363412.
RA   DEMARTINO G.N., ORTH K., MCCULLOUGH M.L., LEE L.W., MUNN T.Z.,
RA   MOOMAW C.R., DAWSON P.A., SLAUGHTER C.A.;
RT   "The primary structures of four subunits of the human,
RT   high-molecular-weight proteinase, macropain (proteasome), are
RT   distinct but homologous.";
RL   BIOCHIM. BIOPHYS. ACTA 1079:29-38(1991).
CC   -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
```

```
CC       WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH
CC       ARG, PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT
CC       NEUTRAL OR SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT
CC       PROTEOLYTIC ACTIVITY.
CC   -!- PATHWAY: IS INVOLVED IN AN ATP/UBIQUITIN-DEPENDENT NON-LYSOSOMAL
CC       PROTEOLYTIC PATHWAY.
CC   -!- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
CC       SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: PROTEASOMES ARE FOUND IN THE CYTOPLASM AND
CC       ALSO IN THE NUCLEUS.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (SHORT AND LONG) ARE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE GENE FOR THIS PROTEIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A; ALSO KNOWN AS THE
CC       PROTEASOME A-TYPE FAMILY. BELONGS TO THE C2 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64992; G190447; ALT_SEQ.
DR   EMBL; D00759; G220022; -.
DR   EMBL; X61969; G296738; -.
DR   PIR; S15897; S15897.
DR   PIR; JC1445; JC1445.
DR   AARHUS/GHENT-2DPAGE; 2223; IEF.
DR   MIM; 602854; -.
DR   PROSITE; PS00388; PROTEASOME_A; 1.
DR   PFAM; PF00227; proteasome; 1.
DR   HSSP; P25156; 1PMA.
KW   PROTEASOME; HYDROLASE; PROTEASE; ACETYLATION; ALTERNATIVE SPLICING.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   VARSPLIC      1      1       M -> MQLSKVK (IN LONG FORM).
SQ   SEQUENCE   263 AA;  29555 MW;  6CD09A93 CRC32;

  Query Match           81.8%;  Score 54;  DB 1;  Length 263;
  Best Local Similarity  75.0%;  Pred. No. 4.20e-01;
  Matches     6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    121 QRYGRRPY 128
          ||| | ||
Qy      1 QRYNRAPY 8


RESULT    4
    GNRP_MOUSE     STANDARD;      PRT;  1262 AA.
    P27671;
    01-AUG-1992 (REL. 23, CREATED)
DT  01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT  01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE  GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP).
GN  RASGRF1 OR CDC25MM OR GRF1.
OS  MUS MUSCULUS (MOUSE).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BALB/C;
RX  MEDLINE; 93010996.
RA  CEN H., LOWY D.D.;
RT  "Isolation of multiple mouse cDNAs with coding homology to
RT  Saccharomyces cerevisiae CDC25: identification of a region related to
RT  Bcr, Vav, Dbl and CDC24.";
RL  EMBO J. 11:4007-4015(1992).
RN  [2]
RP  SEQUENCE OF 791-1262 FROM N.A.
RC  STRAIN=SWISS; TISSUE=BRAIN;
RX  MEDLINE; 92289680.
RA  MARTEGANI E., VANONI M., ZIPPEL R., COCCETTI P., BRAMBILLA R.,
RA  FERRARI C., STURANI E.P., ALBERGHINA L.;
RT  "Cloning by functional complementation of a mouse cDNA encoding a
RT  homologue of CDC25, a Saccharomyces cerevisiae RAS activator.";
RL  EMBO J. 11:2151-2157(1992).
RN  [3]
RP  SEQUENCE OF 1031-1226 FROM N.A.
RX  MEDLINE; 92357779.
RA  WEI W., MOSTELLER R.D., SANYAL P., GONZALES E., MCKINNEY D.,
RA  DASGUPTA C., LI P., LIU B.X., BROEK D.;
RT  "Identification of a mammalian gene structurally and functionally
RT  related to the CDC25 gene of Saccharomyces cerevisiae.";
RL  PROC. NATL. ACAD. SCI. U.S.A. 89:7100-7104(1992).
CC  -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC  -!- TISSUE SPECIFICITY: BRAIN.
CC  -!- SIMILARITY: TO OTHER GUANINE-NUCLEOTIDE RELEASING FACTORS OF THE
CC      CDC25 FAMILY.
CC  -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC  ---------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ---------------------------------------------------------------------
DR  EMBL; L20899; G388247; -.
DR  EMBL; X59868; G50358; -.
DR  PIR; S20730; S20730.
DR  PIR; S22693; S22693.
DR  MGD; MGI:99694; RASGRF1.
DR  PROSITE; PS00720; GDS_CDC25; 1.
DR  PROSITE; PS00741; GDS_CDC24; 1.
DR  PROSITE; PS50003; PH_DOMAIN; 2.
DR  PFAM; PF00169; PH; 2.
DR  PFAM; PF00612; IQ; 1.
DR  PFAM; PF00617; RasGEF; 1.
DR  PFAM; PF00618; RasGEFN; 1.
DR  PFAM; PF00621; RhoGEF; 1.
KW  GUANINE-NUCLEOTIDE RELEASING FACTOR.
FT  DOMAIN       22    130       PH.
FT  DOMAIN      460    588       PH.
FT  CONFLICT   1033   1033       E -> D (IN REF. 3).
SQ  SEQUENCE   1262 AA;  144101 MW;  021C787F CRC32;

  Query Match           77.3%;  Score 51;  DB 1;  Length 1262;
  Best Local Similarity  50.0%;  Pred. No. 1.92e+00;
  Matches     4;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1060 EKYERTPY 1067
          ::|:|:||
Qy      1 QRYNRAPY 8


RESULT    5
ID  FCAR_HUMAN     STANDARD;      PRT;   287 AA.
AC  P24071; Q15728; Q15727; Q13603; Q13604;
DT  01-MAR-1992 (REL. 21, CREATED)
DT  01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT  15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE  IMMUNOGLOBULIN ALPHA FC RECEPTOR PRECURSOR (IGA FC RECEPTOR) (CD89
DE  ANTIGEN).
GN  FCAR OR CD89.
OS  HOMO SAPIENS (HUMAN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 91079769.
RA  MALISZEWSKI C.R., MARCH C.J., SCHOENBORN M.A., GIMPEL S., SHEN L.;
RT  "Expression cloning of a human Fc receptor for IgA.";
RL  J. EXP. MED. 172:1665-1672(1990).
RN  [2]
RP  SEQUENCE FROM N.A.
```

NR

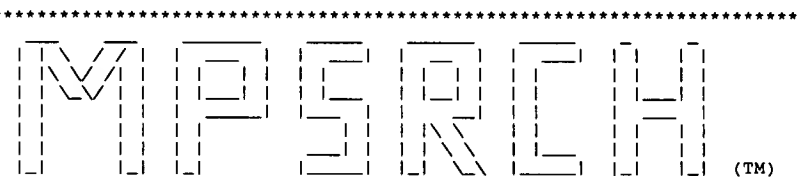
```
RC   TISSUE=BONE MARROW;
RX   MEDLINE; 95363085.
RA   DE WIT T.P.M., MORTON H.C., CAPEL P.J.A., VAN DE WINKEL J.G.J.;
RT   "Structure of the gene for the human myeloid IgA Fc receptor (CD89).";
RL   J. IMMUNOL. 155:1203-1209(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (FORMS A.2 AND A.3).
RC   TISSUE=ALVEOLAR MACROPHAGE, AND MONOCYTES;
RX   MEDLINE; 96247667.
RA   PATRY C., SIBILLE Y., LEHUEN A., MONTEIRO R.C.;
RT   "Identification of Fc alpha receptor (CD89) isoforms generated by
RT   alternative splicing that are differentially expressed between blood
RT   monocytes and alveolar macrophages.";
RL   J. IMMUNOL. 156:4442-4448(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (FORMS B AND B-DELTA-S2).
RA   VAN DIJK T.B., MORTON H.C., CALDENHOVEN E., BRACKE M.,
RA   RAAIJMAKERS J.A.M., LAMMERS J.W.J., KOENDERMAN L., GROOT R.P.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [5]
RP   SEQUENCE FROM N.A. (FORM A.3/RLA2).
RX   MEDLINE; 96433090.
RA   PLEASS R.J., ANDREWS P.D., KERR M.A., WOOF J.M.;
RT   "Alternative splicing of the human IgA Fc receptor CD89 in
RT   neutrophils and eosinophils.";
RL   BIOCHEM. J. 318:771-777(1996).
RN   [6]
RP   SUBUNITS.
RX   MEDLINE; 94375887.
RA   PFEFFERKORN L.C., YEAMAN G.R.;
RT   "Association of IgA-Fc receptors (Fc alpha R) with Fc epsilon RI
RT   gamma 2 subunits in U937 cells. Aggregation induces the tyrosine
RT   phosphorylation of gamma 2.";
RL   J. IMMUNOL. 153:3228-3236(1994).
CC   -!- FUNCTION: BINDS TO THE FC REGION OF IMMUNOGLOBULINS ALPHA.
CC       MEDIATES SEVERAL FUNCTIONS INCLUDING CYTOKINE PRODUCTION.
CC   -!- SUBUNIT: ASSOCIATES WITH THE FC EPSILON RI GAMMA 2 RECEPTOR
CC       INDUCING TYROSINE PHOSPHORYLATION OF GAMMA 2.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (FORM A) AND SOLUBLE
CC       (FORM B).
CC   -!- TISSUE SPECIFICITY: DIFFERENTIALLY EXPRESSED BETWEEN BLOOD AND
CC       MUCOSAL MYELOID CELLS. MONOCYTES EXPRESS ISOFORMS A.1, A.2 AND A.3
CC       WHILE ALVEOLAR MACROPHAGES EXPRESS A.1 AND A.2 TRANSCRIPTS;
CC       HOWEVER THEY EXPRESS ONLY ONE ISOFORM AT THEIR SURFACES.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST FIVE FORMS; A.1 (SHOWN HERE), A.2,
CC       A.3, B AND B-DELTA-S2 ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       2 C2-LIKE DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD89 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd89.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54150; G31330; -.
DR   EMBL; X87767; G1054737; -.
DR   EMBL; X87768; G1054737; JOINED.
DR   EMBL; X87769; G1054737; JOINED.
DR   EMBL; X87766; G1054737; JOINED.
DR   EMBL; X87765; G1054737; JOINED.
DR   EMBL; U43774; G1439556; -.
DR   EMBL; U43677; G1439558; -.
DR   EMBL; U56236; G1326229; -.
DR   EMBL; U56237; G1326231; -.
DR   EMBL; S82919; E302400; -.
DR   PIR; JH0332; JH0332.
DR   MIM; 147045; -.
KW   RECEPTOR; GLYCOPROTEIN; TRANSMEMBRANE; IGA-BINDING PROTEIN;
KW   IMMUNOGLOBULIN FOLD; SIGNAL; ALTERNATIVE SPLICING.
FT   SIGNAL        1    21       POTENTIAL.
FT   CHAIN        22   287       IGA FC RECEPTOR.
FT   DOMAIN       22   227       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    228   246       POTENTIAL.
FT   DOMAIN      247   287       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       42   107       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      139   200       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID     49   100       BY SIMILARITY.
FT   DISULFID    146   193       BY SIMILARITY.
FT   CARBOHYD     65    65       POTENTIAL.
FT   CARBOHYD     79    79       POTENTIAL.
FT   CARBOHYD    141   141       POTENTIAL.
FT   CARBOHYD    177   177       POTENTIAL.
FT   CARBOHYD    186   186       POTENTIAL.
FT   CARBOHYD    198   198       POTENTIAL.
FT   VARSPLIC     12    23       MISSING (IN FORM B-DELTA-S2).
FT   VARSPLIC    121   216       MISSING (IN FORM A.3/RLA2).
FT   VARSPLIC    195   216       MISSING (IN FORM A.2).
FT   VARSPLIC    217   287       DSIHQDYTTQNLIRMAVAGLVLVALLAILVENWHSHTALNK
FT                               EASADVAEPSWSQQMCQPGLTFARTPSVCK -> GRYRPVQ
FT                               PCVWVGCPGPCHRAGI (IN FORM B AND IN FORM
FT                               B-DELTA-S2).
SQ   SEQUENCE   287 AA;  32265 MW;  1B9B01DD CRC32;

  Query Match           75.8%;  Score 50;  DB 1;  Length 287;
  Best Local Similarity  71.4%;  Pred. No. 3.15e+00;
  Matches     5;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    196 WYNRSPY 202
          :|||:||
Qy      2 RYNRAPY 8


RESULT    6
ID   DCAM_SOLTU     STANDARD;      PRT;   360 AA.
AC   Q04694;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME (EC 4.1.1.50) (ADOMETDC)
DE   (SAMDC) (INDUCED STOLEN TIP PROTEIN TUB13).
GN   SAMDC OR TUB13.
OS   SOLANUM TUBEROSUM (POTATO).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC   ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; SOLANUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. RECORD; TISSUE=STOLON TIP;
RX   MEDLINE; 93081725.
RA   TAYLOR M.A., MAD ARIF S.A., KUMAR A., DAVIES H.V., SCOBIE L.A.,
RA   PEARCE S.R., FLAVELL A.J.;
RT   "Expression and sequence analysis of cDNAs induced during the early
RT   stages of tuberisation in different organs of the potato plant
RT   (Solanum tuberosum L.).";
RL   PLANT MOL. BIOL. 20:641-651(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE;
RX   MEDLINE; 95036004.
RA   MAD ARIF S.A., TAYLOR M.A., GEORGE L.A., BUTLER A.R., BURCH L.R.,
RA   DAVIES H.V., STARK M.J., KUMAR A.;
RT   "Characterisation of the S-adenosylmethionine decarboxylase (SAMDC)
RT   gene of potato.";
RL   PLANT MOL. BIOL. 26:327-338(1994).
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYLMETHIONINE = (5-DEOXY-5-ADENOSYL)
CC       (3-AMINOPROPYL)METHYLSULFONIUM SALT + CO(2).
CC   -!- COFACTOR: THIS ENZYME REQUIRES A PYRUVOYL GROUP FOR ITS ACTIVITY.
CC   -!- PATHWAY: DECARBOXYLATION OF S-ADENOSYLMETHIONINE PROVIDES THE
CC       AMINOPROPYL MOIETY REQUIRED FOR SPERMIDINE AND SPERMINE
CC       BIOSYNTHESIS FROM PUTRESCINE.
CC   -!- TISSUE SPECIFICITY: STOLON, ALSO EXPRESSED IN LEAVES, STEMS
```

```
CC       AND ROOTS.
CC   -!- DEVELOPMENTAL STAGE: TRANSCRIBED IN THE STOLON TIP DURING THE
CC       EARLY STAGES OF TUBERIZATION. MAXIMUM EXPRESSION WAS IN NON-
CC       SWELLING STOLON TIPS FROM STAGE B, AND LEVEL DECLINED AS THE
CC       TUBER INCREASED IN SIZE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11680; G21485; -.
DR   EMBL; S74514; G807094; -.
DR   PIR; S28047; S28047.
KW   SPERMIDINE BIOSYNTHESIS; LYASE; DECARBOXYLASE; PYRUVATE; ZYMOGEN.
FT   CHAIN         1     72       S-ADENOSYLMETHIONINE DECARBOXYLASE BETA
FT                                CHAIN (BY SIMILARITY).
FT   CHAIN        73    360       S-ADENOSYLMETHIONINE DECARBOXYLASE ALPHA
FT                                CHAIN (BY SIMILARITY).
FT   SITE         72     73       CLEAVAGE (NONHYDROLYTICAL)
FT                                (BY SIMILARITY).
FT   MOD_RES      73     73       CONVERTED TO A PYRUVOYL GROUP
FT                                (BY SIMILARITY).
FT   ACT_SITE     13     13       IMPORTANT FOR CATALYTIC ACTIVITY (BY
FT                                SIMILARITY).
FT   ACT_SITE     16     16       IMPORTANT FOR CATALYTIC ACTIVITY (BY
FT                                SIMILARITY).
FT   ACT_SITE     87     87       IMPORTANT FOR CATALYTIC ACTIVITY (BY
FT                                SIMILARITY).
FT   CONFLICT    174    174       S -> P (IN REF. 2).
FT   CONFLICT    257    257       T -> S (IN REF. 2).
FT   CONFLICT    291    291       V -> I (IN REF. 2).
FT   CONFLICT    305    305       I -> T (IN REF. 2).
SQ   SEQUENCE   360 AA;  39726 MW;  4F29EFE4 CRC32;

  Query Match           74.2%;  Score 49;  DB 1;  Length 360;
  Best Local Similarity  50.0%;  Pred. No. 5.13e+00;
  Matches     4;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    332 QKFTRTPY 339
          |:: |:||
Qy      1 QRYNRAPY 8


RESULT    7
     RS2_ARATH      STANDARD;      PRT;   285 AA.
     P49688; O22936;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   40S RIBOSOMAL PROTEIN S2.
GN   RPS2 OR T11A07.6.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA   SPRIGGS T.A., MASON T.M., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,
RA   VENTER J.C.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE OF 119-285 FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   RAYNAL M., GRELLET F., LAUDIE M., MEYER Y., COOKE R., DELSENY M.;
RL   SUBMITTED (OCT-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC002339; G2335095; -.
DR   EMBL; Z17622; G16781; -.
DR   PROSITE; PS00585; RIBOSOMAL_S5; 1.
DR   PFAM; PF00333; S5; 1.
DR   HSSP; P02357; 1PKP.
KW   RIBOSOMAL PROTEIN.
FT   CONFLICT    268    269       AS -> ST (IN REF. 2).
FT   CONFLICT    272    281       LSTSKPDPVV -> VSATKVITEG (IN REF. 2).
SQ   SEQUENCE   285 AA;  30878 MW;  4684B4BF CRC32;

  Query Match           72.7%;  Score 48;  DB 1;  Length 285;
  Best Local Similarity  57.1%;  Pred. No. 8.28e+00;
  Matches     4;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    254 RFSRSPY 260
          |::|:||
Qy      2 RYNRAPY 8


RESULT    8
ID   HY14_PIG       STANDARD;      PRT;    14 AA.
AC   P01155;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   21-JUL-1986 (REL. 01, LAST ANNOTATION UPDATE)
DE   HYPOTHALAMIC TETRADECAPEPTIDE.
OS   SUS SCROFA (PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN   [1]
RP   SEQUENCE.
RA   SCHLESINGER D.H., NIALL H.D., LINTHICUM G.L., DUPONT A.,
RA   SCHALLY A.V.;
RL   SUBMITTED (NOV-1976) TO THE PIR DATA BANK.
DR   PIR; A01419; NYPG14.
DR   HSSP; P21856; 1GND.
KW   AMIDATION.
FT   MOD_RES      14     14       AMIDATION.
SQ   SEQUENCE   14 AA;  1648 MW;  59F08C0B CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 14;
  Best Local Similarity  57.1%;  Pred. No. 1.33e+01;
  Matches     4;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db      2 RYGKSPY 8
            || ::||
Qy      2 RYNRAPY 8


RESULT    9
ID   TRA6_NEIME     STANDARD;      PRT;   288 AA.
AC   Q00840;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE   TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS1106 (ORF 1).
OS   NEISSERIA MENINGITIDIS.
OC   BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIACEAE; NEISSERIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B15;
RX   MEDLINE; 92326641.
RA   KNIGHT A.I., NI H., CARTWRIGHT K.A.V., MCFADDEN J.;
RT   "Identification and characterization of a novel insertion sequence,
RT   IS1106, downstream of the porA gene in B15 Neisseria meningitidis.";
```

```
RL   MOL. MICROBIOL. 6:1565-1573(1992).
CC   -!- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION
CC       SEQUENCE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11857; G45066; -.
DR   PIR; S22628; S22628.
KW   TRANSPOSABLE ELEMENT; TRANSPOSITION; DNA-BINDING; DNA RECOMBINATION.
SQ   SEQUENCE   288 AA;  32758 MW;  068FF8FA CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 288;
  Best Local Similarity  71.4%;  Pred. No. 1.33e+01;
  Matches    5;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    250 RYARAAY 256
          || ||:|
Qy      2 RYNRAPY 8


RESULT   10
ID   GDIA_MOUSE     STANDARD;      PRT;   323 AA.
AC   P50396;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   RAB GDP DISSOCIATION INHIBITOR ALPHA (RAB GDI ALPHA) (GDI-1)
DE   (FRAGMENT).
GN   GDI1 OR RABGDIA.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=SKELETAL MUSCLE;
RX   MEDLINE; 94217740.
RA   SHISHEVA A., SUDHOF T.C., CZECH M.P.;
RT   "Cloning, characterization, and expression of a novel GDP
RT   dissociation inhibitor isoform from skeletal muscle.";
RL   MOL. CELL. BIOL. 14:3459-3468(1994).
CC   -!- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC       PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC       SUBSEQUENT BINDING OF GTP TO THEM.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: HIGH EXPRESSION IN BRAIN, LOWER IN OTHER
CC       TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC       INHIBITOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U07950; G516537; -.
DR   MGD; MGI:99846; GDI1.
DR   PFAM; PF00996; GDI; 1.
DR   HSSP; P21856; 1GND.
KW   GTPASE ACTIVATION.
FT   NON_TER       1      1
SQ   SEQUENCE   323 AA;  36601 MW;  EF8EC281 CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 323;
  Best Local Similarity  57.1%;  Pred. No. 1.33e+01;
  Matches    4;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db     94 RYGKSPY 100
          || ::||
Qy      2 RYNRAPY 8


RESULT   11
ID   VIT2_DROME     STANDARD;      PRT;   442 AA.
AC   P02844;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE   VITELLOGENIN II PRECURSOR (YOLK PROTEIN 2).
GN   YP2.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83189120.
RA   HUNG M.-C., WENSINK P.C.;
RT   "Sequence and structure conservation in yolk proteins and their
RT   genes.";
RL   J. MOL. BIOL. 164:481-492(1983).
CC   -!- FUNCTION: VITELLOGENIN IS THE MAJOR YOLK PROTEIN OF EGGS WHERE
CC       IT IS USED AS A FOOD SOURCE DURING EMBRYOGENESIS.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE FAT BODY AND OVARIAN
CC       FOLLICLE CELLS AND ACCUMULATE IN THE OOCYTE.
CC   -!- INDUCTION: IN MALES BY BETA-ECDYSONE.
CC   -!- SIMILARITY: PARTIAL, TO LIPASES. STRONG TO OTHER VITTELOGENINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J01157; -; NOT_ANNOTATED_CDS.
DR   PIR; A03333; VJFF2.
DR   FLYBASE; FBgn0005391; Yp2.
DR   PFAM; PF00151; lipase; 1.
KW   YOLK; SIGNAL.
FT   SIGNAL        1     20
FT   CHAIN        21    442       VITELLOGENIN II.
SQ   SEQUENCE   442 AA;  49678 MW;  A433872E CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 442;
  Best Local Similarity  75.0%;  Pred. No. 1.33e+01;
  Matches    6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    160 QRYNLQPY 167
          |||| ||
Qy      1 QRYNRAPY 8


RESULT   12
ID   GDIB_HUMAN     STANDARD;      PRT;   445 AA.
AC   P50395;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   RAB GDP DISSOCIATION INHIBITOR BETA (RAB GDI BETA) (GDI-2).
GN   GDI2 OR RABGDIB.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
```

```
RA   ASADA M., KAIBUCHI K., TAKAI Y.;
RL   SUBMITTED (APR-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   TISSUE SPECIFICITY.
RX   MEDLINE; 95359978.
RA   BACHNER D., SEDLACEK Z., KORN B., HAMEISTER H., POUSTKA A.;
RT   "Expression patterns of two human genes coding for different rab GDP-
RT   dissociation inhibitors (GDIs), extremely conserved proteins involved
RT   in cellular transport.";
RL   HUM. MOL. GENET. 4:701-708(1995).
CC   -!- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC       PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC       SUBSEQUENT BINDING OF GTP TO THEM.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   -!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC       INHIBITOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13988; G285975; -.
DR   MIM; 600767; -.
DR   PFAM; PF00996; GDI; 1.
DR   HSSP; P21856; 1GND.
KW   GTPASE ACTIVATION.
SQ   SEQUENCE   445 AA;  50664 MW;  858E540A CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 445;
  Best Local Similarity  57.1%;  Pred. No. 1.33e+01;
  Matches     4;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    218 RYGKSPY 224
          || ::||
Qy      2 RYNRAPY 8


RESULT   13
ID   GDIB_RAT       STANDARD;      PRT;   445 AA.
AC   P50399;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   RAB GDP DISSOCIATION INHIBITOR BETA (RAB GDI BETA) (GDI-2).
GN   GDI2 OR RABGDIB.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 94245743.
RA   NISHIMURA N., NAKAMURA H., TAKAI Y., SANO K.;
RT   "Molecular cloning and characterization of two rab GDI species from
RT   rat brain: brain-specific and ubiquitous types.";
RL   J. BIOL. CHEM. 269:14191-14198(1994).
RN   [2]
RP   SEQUENCE OF 30-54 AND 58-74, AND CHARACTERIZATION.
RX   MEDLINE; 95298038.
RA   ARAKI K., NAKANISHI H., HIRANO H., KATO M., SASAKI T., TAKAI Y.;
RT   "Purification and characterization of Rab GDI beta from rat brain.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 211:296-305(1995).
CC   -!- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC       PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC       SUBSEQUENT BINDING OF GTP TO THEM.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-ASSOCIATED.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC   -!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC       INHIBITOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74401; G396433; -.
DR   PFAM; PF00996; GDI; 1.
DR   HSSP; P21856; 1GND.
KW   GTPASE ACTIVATION.
SQ   SEQUENCE   445 AA;  50685 MW;  842F64F7 CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 445;
  Best Local Similarity  57.1%;  Pred. No. 1.33e+01;
  Matches     4;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    218 RYGKSPY 224
          || ::||
Qy      2 RYNRAPY 8


RESULT   14
ID   GDIB_MOUSE     STANDARD;      PRT;   445 AA.
AC   P50397;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   RAB GDP DISSOCIATION INHIBITOR BETA (RAB GDI BETA) (GDI-2).
GN   GDI2 OR RABGDIB.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=SKELETAL MUSCLE;
RX   MEDLINE; 94217740.
RA   SHISHEVA A., SUDHOF T.C., CZECH M.P.;
RT   "Cloning, characterization, and expression of a novel GDP
RT   dissociation inhibitor isoform from skeletal muscle.";
RL   MOL. CELL. BIOL. 14:3459-3468(1994).
CC   -!- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC       PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC       SUBSEQUENT BINDING OF GTP TO THEM.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-ASSOCIATED.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC   -!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC       INHIBITOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U07951; G1486480; ALT_SEQ.
DR   MGD; MGI:99845; GDI2.
DR   PFAM; PF00996; GDI; 1.
DR   HSSP; P21856; 1GND.
KW   GTPASE ACTIVATION.
SQ   SEQUENCE   445 AA;  50653 MW;  79A76704 CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 445;
  Best Local Similarity  57.1%;  Pred. No. 1.33e+01;
  Matches     4;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    218 RYGKSPY 224
          || ::||
Qy      2 RYNRAPY 8
```

```
RESULT   15
ID   GDIA_HUMAN     STANDARD;      PRT;   447 AA.
AC   P31150; P50394;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   RAB GDP DISSOCIATION INHIBITOR ALPHA (RAB GDI ALPHA) (GDI-1) (XAP-4).
GN   GDI1 OR RABGDIA OR XAP4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95152170.
RA   SEDLACEK Z., KONECKI D.S., KORN B., KLAUCK S.M., POUSTKA A.;
RT   "Evolutionary conservation and genomic organization of XAP-4, an Xq28
RT   located gene coding for a human rab GDP-dissociation inhibitor
     (GDI).";
     MAMM. GENOME 5:633-639(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE; 96062207.
RA   NISHIMURA N., GOJI J., NAKAMURA H., ORITA S., TAKAI Y., SANO K.;
RT   "Cloning of a brain-type isoform of human Rab GDI and its expression
RT   in human neuroblastoma cell lines and tumor specimens.";
RL   CANCER RES. 55:5445-5450(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   CHEN E.Y., ZOLLO M., MAZZARELLA R.A., CICCODICOLA A., CHEN C.N.,
RA   ZUO L., HEINER C., BUROUGH F.W., RIPETTO M., SCHLESSINGER D.,
RA   D'URSO M.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE OF 143-181 FROM N.A.
RA   HOCHGESCHWENDER U.;
RL   SUBMITTED (SEP-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [5]
RP   SEQUENCE OF 328-436 FROM N.A.
RA   BHAT K.S.;
RL   SUBMITTED (XXX-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [6]
RP   SEQUENCE OF 349-361.
RC   TISSUE=KERATINOCYTES;
RX   MEDLINE; 93162043.
     RASMUSSEN H.H., VAN DAMME J., PUYPE M., GESSER B., CELIS J.E.,
     VANDEKERCKHOVE J.;
RT   "Microsequences of 145 proteins recorded in the two-dimensional gel
RT   protein database of normal human epidermal keratinocytes.";
RL   ELECTROPHORESIS 13:960-969(1992).
RN   [7]
RP   TISSUE SPECIFICITY.
RX   MEDLINE; 95359978.
RA   BACHNER D., SEDLACEK Z., KORN B., HAMEISTER H., POUSTKA A.;
RT   "Expression patterns of two human genes coding for different rab GDP-
RT   dissociation inhibitors (GDIs), extremely conserved proteins involved
RT   in cellular transport.";
RL   HUM. MOL. GENET. 4:701-708(1995).
CC   -!- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC       PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC       SUBSEQUENT BINDING OF GTP TO THEM.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: BRAIN; PREDOMINANT IN NEURAL AND SENSORY
CC       TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC       INHIBITOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79354; G695523; -.
DR   EMBL; X79355; G695523; JOINED.
DR   EMBL; X79356; G695523; JOINED.
DR   EMBL; X79357; G695523; JOINED.
DR   EMBL; X79358; G695523; JOINED.
DR   EMBL; X79359; G695523; JOINED.
DR   EMBL; X79360; G695523; JOINED.
DR   EMBL; X79364; G695523; JOINED.
DR   EMBL; X79361; G695523; JOINED.
DR   EMBL; X79362; G695523; JOINED.
DR   EMBL; X79363; G695523; JOINED.
DR   EMBL; X79353; G695585; -.
DR   EMBL; L44140; G1203973; -.
DR   EMBL; U14623; G540251; -.
DR   EMBL; D45021; G624873; -.
DR   EMBL; L05086; G553301; -.
DR   AARHUS/GHENT-2DPAGE; 8408; IEF.
DR   MIM; 300104; -.
DR   PFAM; PF00996; GDI; 1.
DR   HSSP; P21856; 1GND.
KW   GTPASE ACTIVATION.
FT   CONFLICT     34     34       D -> G (IN REF. 2).
FT   CONFLICT     36     36       N -> K (IN REF. 2).
FT   CONFLICT    149    151       NFD -> GTY (IN REF. 2).
FT   CONFLICT    347    347       G -> V (IN REF. 4).
FT   CONFLICT    409    409       H -> Q (IN REF. 4).
FT   CONFLICT    416    416       D -> G (IN REF. 4).
FT   CONFLICT    442    442       F -> S (IN REF. 2).
SQ   SEQUENCE   447 AA;  50582 MW;  6E84EC0D CRC32;

  Query Match           71.2%;  Score 47;  DB 1;  Length 447;
  Best Local Similarity  57.1%;  Pred. No. 1.33e+01;
  Matches     4;  Conservative     2;  Mismatches   1;  Indels    0;  Gaps    0;

Db    218 RYGKSPY 224
          || ::||
Qy      2 RYNRAPY 8


Search completed: Thu Sep  2 11:18:09 1999
Job time : 8 secs.
```



MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 11:19:02 1999; MasPar time 1.35 Seconds
67.852 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-3
Description: (1-9) from US08599226.pep
Perfect Score: 66
Sequence: 1 QRYNRAPYX 9

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 15.499; Variance 48.116; scale 0.322

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	54	81.8	269	1	US-08-202-	Sequence 4, Applicatio	1.41e+01
2	50	75.8	287	4	5198342-2	Patent No. 5198342.	3.83e+01
3	50	75.8	287	1	US-07-971-	Sequence 2, Applicatio	3.83e+01
4	45	68.2	246	2	US-08-604-	Sequence 3, Applicatio	1.30e+02
5	45	68.2	466	2	US-08-604-	Sequence 4, Applicatio	1.30e+02
6	45	68.2	507	3	PCT-US95-0	Sequence 2, Applicatio	1.30e+02
7	45	68.2	507	2	US-08-604-	Sequence 5, Applicatio	1.30e+02
8	44	66.7	33	2	US-08-724-	Sequence 12, Applicati	1.65e+02
9	44	66.7	218	2	US-08-343-	Sequence 20, Applicati	1.65e+02
10	44	66.7	400	1	US-07-989-	Sequence 2, Applicatio	1.65e+02
11	44	66.7	454	2	US-08-748-	Sequence 1, Applicatio	1.65e+02
12	44	66.7	529	2	US-08-467-	Sequence 2, Applicatio	1.65e+02
13	44	66.7	529	3	PCT-US94-0	Sequence 2, Applicatio	1.65e+02
14	44	66.7	1376	3	PCT-US95-1	Sequence 3, Applicatio	1.65e+02
15	44	66.7	1376	2	US-08-420-	Sequence 3, Applicatio	1.65e+02
16	43	65.2	107	3	PCT-US95-0	Sequence 87, Applicati	2.09e+02
17	43	65.2	107	2	US-08-899-	Sequence 87, Applicati	2.09e+02
18	43	65.2	107	1	US-08-276-	Sequence 87, Applicati	2.09e+02
19	43	65.2	107	2	US-08-899-	Sequence 87, Applicati	2.09e+02
20	43	65.2	387	1	US-07-996-	Sequence 2, Applicatio	2.09e+02
21	43	65.2	387	2	US-08-446-	Sequence 2, Applicatio	2.09e+02
22	43	65.2	387	3	PCT-US93-1	Sequence 2, Applicatio	2.09e+02
23	43	65.2	406	1	US-07-996-	Sequence 4, Applicatio	2.09e+02
24	43	65.2	406	2	US-08-446-	Sequence 4, Applicatio	2.09e+02
25	43	65.2	406	3	PCT-US93-1	Sequence 4, Applicatio	2.09e+02
26	43	65.2	718	1	US-08-445-	Sequence 4, Applicatio	2.09e+02
27	43	65.2	718	1	US-08-444-	Sequence 4, Applicatio	2.09e+02
28	42	63.6	52	2	US-08-162-	Sequence 43, Applicati	2.65e+02
29	42	63.6	109	3	PCT-US93-0	Sequence 23, Applicati	2.65e+02
30	42	63.6	109	2	US-08-162-	Sequence 23, Applicati	2.65e+02
31	42	63.6	150	2	US-08-592-	Sequence 14, Applicati	2.65e+02
32	42	63.6	253	2	US-08-592-	Sequence 4, Applicatio	2.65e+02
33	42	63.6	255	2	US-08-592-	Sequence 6, Applicatio	2.65e+02
34	42	63.6	255	2	US-08-592-	Sequence 10, Applicati	2.65e+02
35	42	63.6	255	2	US-08-576-	Sequence 2, Applicatio	2.65e+02
36	42	63.6	273	2	US-08-592-	Sequence 8, Applicatio	2.65e+02
37	42	63.6	300	2	US-08-592-	Sequence 2, Applicatio	2.65e+02
38	42	63.6	450	2	US-08-712-	Sequence 1, Applicatio	2.65e+02
39	42	63.6	1708	1	US-08-493-	Sequence 2, Applicatio	2.65e+02
40	42	63.6	1708	2	US-08-508-	Sequence 2, Applicatio	2.65e+02
41	42	63.6	3056	2	US-08-508-	Sequence 8, Applicatio	2.65e+02
42	42	63.6	3056	2	US-08-629-	Sequence 3, Applicatio	2.65e+02
43	41	62.1	12	3	PCT-US96-0	Sequence 143, Applicat	3.35e+02
44	41	62.1	109	3	PCT-US93-0	Sequence 3, Applicatio	3.35e+02
45	41	62.1	449	1	US-07-917-	Sequence 2, Applicatio	3.35e+02

ALIGNMENTS

```
RESULT    1
ID   US-08-202-857-4     STANDARD;     PRT;   269 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08202857
XX
CC    Sequence 4, Application US/08202857
CC    Patent No. 5635345
CC     GENERAL INFORMATION:
CC       APPLICANT:  Scherrer, Klaus
CC       APPLICANT:  Bureau, Jean-Paul
CC       APPLICANT:  Bey, Fay al
CC       TITLE OF INVENTION:  Diagnostic Method
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PRO-SOMA SARL C/O AKZO PHARMA
CC         STREET:  1330-A PICCARD DRIVE
CC         CITY:  ROCKVILLE
CC         STATE:  MARYLAND
CC         COUNTRY:  USA
CC         ZIP:  20850-4377
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PS/2
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  Microsoft Word
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/202,857
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/860,492
CC         FILING DATE:  11-JUNE-1992
CC         APPLICATION NUMBER:  PCT/EP91/01945
CC         FILING DATE:  10-OCT-1991
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  BOBROWICZ, DONNA
CC         REGISTRATION NUMBER:  32,196
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  301-258-5200
CC         TELEFAX:  301-977-1403
CC      INFORMATION FOR SEQ ID NO:  4:
CC        SEQUENCE CHARACTERISTICS:
CC          LENGTH:  269 amino acids
```

```
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       ORIGINAL SOURCE:
CC         ORGANISM:  human
SQ   SEQUENCE   269 AA; 30227 MW; 341084 CN;

  Query Match           81.8%;  Score 54;  DB 1;  Length 269;
  Best Local Similarity  75.0%;  Pred. No. 1.41e+01;
  Matches     6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    127 QRYGRRPY 134
          ||| | ||
Qy      1 QRYNRAPY 8


RESULT    2
ID   5198342-2     STANDARD;     PRT;   311 AA.
XX
AC   xxxxxx
XX
DT   01-JAN-1900
XX
DE   Patent No. 5198342.
XX
CC   Patent No. 5198342
CC       APPLICANT: MALIISZEWSKI,CHARLES R.
CC       TITLE OF INVENTION: DNA ENCOGING IGA FC RECEPTORS
CC       NUMBER OF SEQUENCES: 9
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/548,059
CC         FILING DATE: 05-JUL-1990
CC   SEQ ID NO:2:
CC         LENGTH: 287
SQ   SEQUENCE   311 AA; 34908 MW; 558002 CN;

  Query Match           75.8%;  Score 50;  DB 4;  Length 287;
  Best Local Similarity  71.4%;  Pred. No. 3.83e+01;
  Matches     5;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    196 WYNRSPY 202
          :|||:||
Qy      2 RYNRAPY 8


RESULT    3
ID   US-07-971-092-2     STANDARD;     PRT;   287 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/07971092
XX
CC    Sequence 2, Application US/07971092
CC    Patent No. 5328987
CC     GENERAL INFORMATION:
CC       APPLICANT:  Maliszewski, Charles R.
CC       TITLE OF INVENTION:  HuIgA Fc Receptor
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Immunex
CC         STREET:  51 University
CC         CITY:  Seattle
CC         STATE:  WA
CC         COUNTRY:  USA
CC         ZIP:  98101
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/971,092
CC         FILING DATE:  19921104
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Perkins, Patricia A.
CC         REGISTRATION NUMBER:  34693
CC         REFERENCE/DOCKET NUMBER:  2603
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  287 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   287 AA; 32265 MW; 435416 CN;

  Query Match           75.8%;  Score 50;  DB 1;  Length 287;
  Best Local Similarity  71.4%;  Pred. No. 3.83e+01;
  Matches     5;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    196 WYNRSPY 202
          :|||:||
Qy      2 RYNRAPY 8


RESULT    4
ID   US-08-604-989A-3     STANDARD;     PRT;   246 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08604989A
XX
CC    Sequence 3, Application US/08604989A
CC    Patent No. 5834208
CC     GENERAL INFORMATION:
CC       APPLICANT:  Sakano, S.
CC       TITLE OF INVENTION:  No. 5834208el Tyrosine Kinase
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds LLP
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/604,989A
CC         FILING DATE:  February 23, 1996
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Charles E. Miller
CC         REGISTRATION NUMBER:  24,576
CC         REFERENCE/DOCKET NUMBER:  1920-026
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-8864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  246 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       ORIGINAL SOURCE:
CC         ORGANISM:  human
CC         STRAIN:  UT-7
SQ   SEQUENCE   246 AA; 27425 MW; 326654 CN;
```

```
  Query Match           68.2%;  Score 45;  DB 2;  Length 246;
  Best Local Similarity  83.3%;  Pred. No. 1.30e+02;
  Matches     5;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    188 YGRAPY 193
          | ||||
Qy      3 YNRAPY 8


RESULT    5
ID   US-08-604-989A-4     STANDARD;     PRT;   466 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08604989A
XX
CC    Sequence 4, Application US/08604989A
CC    Patent No. 5834208
CC     GENERAL INFORMATION:
CC       APPLICANT:  Sakano, S.
CC       TITLE OF INVENTION:  No. 5834208el Tyrosine Kinase
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds LLP
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/604,989A
CC         FILING DATE:  February 23, 1996
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Charles E. Miller
CC         REGISTRATION NUMBER:  24,576
CC         REFERENCE/DOCKET NUMBER:  1920-026
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-8864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  466 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       ORIGINAL SOURCE:
CC         ORGANISM:  human
CC         STRAIN:  UT-7
SQ   SEQUENCE   466 AA; 51898 MW; 1114727 CN;

  Query Match           68.2%;  Score 45;  DB 2;  Length 466;
  Best Local Similarity  83.3%;  Pred. No. 1.30e+02;
  Matches     5;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    379 YGRAPY 384
          | ||||
Qy      3 YNRAPY 8


RESULT    6
ID   PCT-US95-05008-2     STANDARD;     PRT;   507 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9505008
XX
CC    Sequence 2, Application PC/TUS9505008
CC     GENERAL INFORMATION:
CC       APPLICANT:  Sugen, Inc.
CC       APPLICANT:  515 Galveston Drive
CC       APPLICANT:  Redwood City, California 94063-4720
CC       APPLICANT:  United States of America
CC       APPLICANT:  Wissenschaften E.V.
CC       APPLICANT:  Hofgarten Str. 2
CC       APPLICANT:  Munchen 80539
CC       APPLICANT:  Germany
CC       TITLE OF INVENTION:  Novel Megakaryocytic Protein Tyrosine
CC       TITLE OF INVENTION:  Kinases
CC       NUMBER OF SEQUENCES:  21
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/05008
CC         FILING DATE:  24-APR-1995
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/232,545
CC         FILING DATE:  22-APR-1994
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Coruzzi, Laura A.
CC         REGISTRATION NUMBER:  30,742
CC         REFERENCE/DOCKET NUMBER:  7683-074
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212)790-9090
CC         TELEFAX:  (212)869-9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  507 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   507 AA; 56469 MW; 1316735 CN;

  Query Match           68.2%;  Score 45;  DB 3;  Length 507;
  Best Local Similarity  83.3%;  Pred. No. 1.30e+02;
  Matches     5;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    420 YGRAPY 425
          | ||||
Qy      3 YNRAPY 8


RESULT    7
ID   US-08-604-989A-5     STANDARD;     PRT;   507 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/08604989A
XX
```

```
CC    Sequence 5, Application US/08604989A
CC    Patent No. 5834208
CC     GENERAL INFORMATION:
CC       APPLICANT:  Sakano, S.
CC       TITLE OF INVENTION:  No. 5834208el Tyrosine Kinase
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds LLP
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/604,989A
CC         FILING DATE:  February 23, 1996
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Charles E. Miller
CC         REGISTRATION NUMBER:  24,576
CC         REFERENCE/DOCKET NUMBER:  1920-026
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-8864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  507 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       ORIGINAL SOURCE:
CC         ORGANISM:  human
CC         STRAIN:  UT-7
SQ   SEQUENCE   507 AA; 56491 MW; 1317560 CN;

  Query Match           68.2%;  Score 45;  DB 2;  Length 507;
  Best Local Similarity  83.3%;  Pred. No. 1.30e+02;
  Matches     5;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    420 YGRAPY 425
          | ||||
Qy      3 YNRAPY 8


RESULT    8
ID   US-08-724-194-12     STANDARD;     PRT;   33 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 12, Application US/08724194
XX
CC    Sequence 12, Application US/08724194
CC    Patent No. 5824875
CC     GENERAL INFORMATION:
CC       APPLICANT:  RANU, RAJINDER S.
CC       TITLE OF INVENTION:  ONE-AMINOCYCLOPROPANE-1-CARBOXYLATE
CC       TITLE OF INVENTION:  SYNTHASE GENES FROM PELARGONIUM TO CONTROL ETHYLENE LEVELS
CC       TITLE OF INVENTION:  IN GERANIUMS
CC       NUMBER OF SEQUENCES:  13
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  SANTANGELO LAW OFFICES PC
CC         STREET:  315 WEST OAK STREET, STE 701
CC         CITY:  FORT COLLINS
CC         STATE:  CO
CC         COUNTRY:  USA
CC         ZIP:  80521
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/724,194
CC         FILING DATE:  01-OCT-1996
CC         CLASSIFICATION:  800
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  SANTANGELO, LUKE
CC         REGISTRATION NUMBER:  31,997
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (970) 224-3100
CC     INFORMATION FOR SEQ ID NO:  12:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  33 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   33 AA; 3822 MW; 5035 CN;

  Query Match           66.7%;  Score 44;  DB 2;  Length 33;
  Best Local Similarity  66.7%;  Pred. No. 1.65e+02;
  Matches     4;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db      8 YDRDPY 13
          |:| ||
Qy      3 YNRAPY 8


RESULT    9
ID   US-08-343-101A-20     STANDARD;     PRT;   218 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 20, Application US/08343101A
XX
CC    Sequence 20, Application US/08343101A
CC    Patent No. 5830759
CC     GENERAL INFORMATION:
CC       APPLICANT:  Chang, Yuan
CC       APPLICANT:  Moore, Patrick S.
CC       TITLE OF INVENTION:  Unique Associated Kaposi's Sarcoma
CC       TITLE OF INVENTION:  Virus Sequences And Uses Thereof
CC       NUMBER OF SEQUENCES:  22
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham
CC         STREET:  1185 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.24
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/343,101A
CC         FILING DATE:
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White Esq., John P.
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  45185-A
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-278-0400
CC         TELEFAX:  212-391-0526
```

```
CC     INFORMATION FOR SEQ ID NO:  20:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  218 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  N
CC       ANTI-SENSE:  N
CC       FEATURE:
CC         NAME/KEY:  Peptide
CC         LOCATION:  1..218
CC         OTHER INFORMATION:
SQ   SEQUENCE   218 AA; 24345 MW; 253377 CN;

  Query Match           66.7%;  Score 44;  DB 2;  Length 218;
  Best Local Similarity  71.4%;  Pred. No. 1.65e+02;
  Matches     5;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    119 RYNRGLY 125
          |||| : |
Qy      2 RYNRAPY 8


RESULT   10
ID   US-07-989-991A-2     STANDARD;     PRT;   400 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/07989991A
XX
CC    Sequence 2, Application US/07989991A
CC    Patent No. 5525502
CC     GENERAL INFORMATION:
CC       APPLICANT:  Thireos, George
CC       APPLICANT:  Kafetzopoulos, Dimitris
CC       TITLE OF INVENTION:  DNA ENCODING CHITIN DEACETYLASE
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
CC         STREET:  Two Militia Drive
CC         CITY:  Lexington
CC         STATE:  MA
CC         COUNTRY:  USA
CC         ZIP:  02173
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/989,991A
CC         FILING DATE:  07-DEC-1992
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Brook, David E.
CC         REGISTRATION NUMBER:  22,592
CC         REFERENCE/DOCKET NUMBER:  BTT91-01A
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  617 861-6240
CC         TELEFAX:  617 861-9540
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  400 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   400 AA; 43857 MW; 860821 CN;

  Query Match           66.7%;  Score 44;  DB 1;  Length 400;
  Best Local Similarity  71.4%;  Pred. No. 1.65e+02;
  Matches     5;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    228 RYWRPPY 234
          || |:||
Qy      2 RYNRAPY 8


RESULT   11
ID   US-08-748-485-1     STANDARD;     PRT;   454 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08748485
XX
CC    Sequence 1, Application US/08748485
CC    Patent No. 5817480
CC     GENERAL INFORMATION:
CC       APPLICANT:  Au-Young, Janice
CC       APPLICANT:  Guegler, Karl J.
CC       APPLICANT:  Goli, Surya K.
CC       APPLICANT:  Murry, Lynn E.
CC       TITLE OF INVENTION:  NOVEL HISTAMINE H2 RECEPTOR
CC       NUMBER OF SEQUENCES:  8
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  US
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/748,485
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0159 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  454 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  Consensus
CC         CLONE:  1722180
SQ   SEQUENCE   454 AA; 50589 MW; 1146364 CN;

  Query Match           66.7%;  Score 44;  DB 2;  Length 454;
  Best Local Similarity  50.0%;  Pred. No. 1.65e+02;
  Matches     4;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    343 DRYYREPF 350
          :|| | |:
Qy      1 QRYNRAPY 8


RESULT   12
```

ID US-08-467-568-2 STANDARD; PRT; 529 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/08467568
XX
CC Sequence 2, Application US/08467568
CC Patent No. 5817477
CC GENERAL INFORMATION:
CC APPLICANT: SOPPET, DANIEL R
CC TITLE OF INVENTION: ADRENERGIC RECEPTOR
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
CC ADDRESSEE: Stewart & Olstein
CC STREET: 6 Becker Farm Road
CC CITY: Roseland
CC STATE: NJ
CC COUNTRY: USA
CC ZIP: 07068-1739
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/467,568
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Ferraro, Gregory D
CC REGISTRATION NUMBER: 36,134
CC REFERENCE/DOCKET NUMBER: 325800-324
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-994-1700
CC TELEFAX: 201-994-1744
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 529 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 529 AA; 58358 MW; 1481658 CN;

Query Match 66.7%; Score 44; DB 2; Length 529;
Best Local Similarity 50.0%; Pred. No. 1.65e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 342 DRYYREPF 349
:|| | |:
Qy 1 QRYNRAPY 8

RESULT 13
ID PCT-US94-09051-2 STANDARD; PRT; 529 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application PC/TUS9409051
XX
CC Sequence 2, Application PC/TUS9409051
CC GENERAL INFORMATION:
CC APPLICANT: LI, ET AL.
CC TITLE OF INVENTION: Adrenergic Receptor
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
CC ADDRESSEE: CECCHI, STEWART & OLSTEIN
CC STREET: 6 BECKER FARM ROAD
CC CITY: ROSELAND
CC STATE: NEW JERSEY
CC COUNTRY: USA
CC ZIP: 07068
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5 INCH DISKETTE
CC COMPUTER: IBM PS/2
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: WORD PERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/09051
CC FILING DATE: Submitted herewith
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: FERRARO, GREGORY D.
CC REGISTRATION NUMBER: 36,134
CC REFERENCE/DOCKET NUMBER: 325800-194
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-994-1700
CC TELEFAX: 201-994-1744
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 529 AMINO ACIDS
CC TYPE: AMINO ACID
CC STRANDEDNESS:
CC TOPOLOGY: LINEAR
CC MOLECULE TYPE: PROTEIN
SQ SEQUENCE 529 AA; 58358 MW; 1481658 CN;

Query Match 66.7%; Score 44; DB 3; Length 529;
Best Local Similarity 50.0%; Pred. No. 1.65e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 342 DRYYREPF 349
:|| | |:
Qy 1 QRYNRAPY 8

RESULT 14
ID PCT-US95-10194-3 STANDARD; PRT; 1376 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 3, Application PC/TUS9510194
XX
CC Sequence 3, Application PC/TUS9510194
CC GENERAL INFORMATION:
CC APPLICANT: The Trustees of Columbia University in the City of New York
CC APPLICANT: City
CC TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
CC TITLE OF INVENTION: SEQUENCES AND USES THEREOF
CC NUMBER OF SEQUENCES: 45
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooper & Dunham LLP
CC STREET: 1185 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10194
CC FILING DATE:
CC CLASSIFICATION:

CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P.
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 278-0400
CC TELEFAX: (212) 391-0525
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1376 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1376 AA; 153401 MW; 9818578 CN;

Query Match 66.7%; Score 44; DB 3; Length 1376;
Best Local Similarity 71.4%; Pred. No. 1.65e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1277 RYNRGLY 1283
|||| : |
2 RYNRAPY 8

RESULT 15
ID US-08-420-235B-3 STANDARD; PRT; 1376 AA.
XX
AC XXXXXX
XX
DT
XX
DE Sequence 3, Application US/08420235B
XX
CC Sequence 3, Application US/08420235B
CC Patent No. 5801042
CC GENERAL INFORMATION:
CC APPLICANT: Chang, Yuan
CC APPLICANT: Moore, Patrick S.
CC TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
CC TITLE OF INVENTION: SEQUENCES AND USES THEREOF
CC NUMBER OF SEQUENCES: 47
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooper & Dunham LLP
CC STREET: 1185 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/420,235B
CC FILING DATE:
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P.
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 45185-B
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 278-0400
CC TELEFAX: (212) 391-0525
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1376 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1376 AA; 153401 MW; 9818578 CN;

Query Match 66.7%; Score 44; DB 2; Length 1376;
Best Local Similarity 71.4%; Pred. No. 1.65e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1277 RYNRGLY 1283
|||| : |
Qy 2 RYNRAPY 8

Search completed: Thu Sep 2 11:19:10 1999
Job time : 8 secs.

```
*****************************************************************************
                           MPSRCH (TM)
*****************************************************************************

      Release 3.1A John F. Collins, Biocomputing Research Unit.
      Copyright (c) 1993-1998  University of Edinburgh, U.K.
            Distribution rights by Oxford Molecular Ltd

MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

Run on:          Thu Sep  2 11:17:30 1999;  MasPar time 3.08 Seconds
                                            116.921 Million cell updates/sec
...ular output not generated.

Title:           >US-08-599-226-3
Description:     (1-9) from US08599226.pep
Perfect Score:   66
Sequence:        1 QRYNRAPYX 9

Scoring table:   PAM 150
                 Gap 15

Searched:        122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:        pir60
                   1:pir1 2:pir2 3:pir3 4:pir4

Statistics:      Mean 22.973;  Variance 30.279;  scale 0.759

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	57	86.4	279	1	SNFF5K	multicatalytic endope	3.45e-01
2	54	81.8	263	1	SNRTC2	multicatalytic endope	1.43e+00
3	54	81.8	269	2	JC1445	multicatalytic endope	1.43e+00
4	53	80.3	790	2	S77032	ABC-type transport pr	2.27e+00
5	52	78.8	321	2	B71854	type II DNA modificat	3.58e+00
6	51	77.3	1260	2	S28407	guanine nucleotide-ex	5.63e+00
7	50	75.8	239	2	G02630	FcalphaRb - human	8.80e+00
8	50	75.8	287	2	JH0332	IgA (Fc) receptor, my	8.80e+00
9	50	75.8	521	2	I39956	neutral proteinase (E	8.80e+00
10	49	74.2	360	2	S52662	S-adenosylmethionine	1.37e+01
11	49	74.2	360	2	S28047	TUB13 protein - potat	1.37e+01
12	48	72.7	367	2	S49009	fork head protein 2 -	2.11e+01
13	48	72.7	370	2	S49008	fork head protein - A	2.11e+01
14	48	72.7	534	2	S71800	transcription factor	2.11e+01
15	48	72.7	1032	2	S74487	hypothetical protein	2.11e+01
16	47	71.2	14	1	NYPG14	hypothalamic tetradec	3.24e+01
17	47	71.2	288	2	S22628	hypothetical protein	3.24e+01
18	47	71.2	323	2	C56024	GDP dissociation inhi	3.24e+01
19	47	71.2	355	2	S76940	hypothetical protein	3.24e+01
20	47	71.2	440	2	F70792	hypothetical protein	3.24e+01
21	47	71.2	442	1	VJFF2	vitellogenin II precu	3.24e+01
22	47	71.2	445	2	C56956	GDP dissociation inhi	3.24e+01
23	47	71.2	445	2	B54091	rab GDP dissociation	3.24e+01
24	47	71.2	445	2	A56024	GDP dissociation inhi	3.24e+01
25	47	71.2	447	2	A54091	rab GDP dissociation	3.24e+01
26	47	71.2	447	2	A35652	smg p25A regulatory p	3.24e+01
27	47	71.2	447	2	I37082	GDP-dissociation inhi	3.24e+01
28	47	71.2	447	2	B56024	GDP dissociation inhi	3.24e+01
29	47	71.2	448	2	S36746	GDP dissociation inhi	3.24e+01
30	47	71.2	451	2	S44446	GDP dissociation inhi	3.24e+01
31	47	71.2	515	2	T03070	hypothetical protein	3.24e+01
32	47	71.2	520	2	B71143	hypothetical protein	3.24e+01
33	47	71.2	541	2	S01957	hypothetical protein	3.24e+01
34	47	71.2	689	2	S45901	probable membrane pro	3.24e+01
35	47	71.2	722	2	C71411	hypothetical protein	3.24e+01
36	47	71.2	1320	2	S66279	proline dehydrogenase	3.24e+01
37	47	71.2	1320	2	D64843	proline dehydrogenase	3.24e+01
38	46	69.7	106	2	G02071	helicase - human (fra	4.93e+01
39	46	69.7	182	2	S45202	hypoxanthine phosphor	4.93e+01
40	46	69.7	365	2	C37753	quinolinate synthase	4.93e+01
41	46	69.7	419	3	JE0389	catabolite repressor	4.93e+01
42	46	69.7	613	2	S55615	thymidine kinase (EC	4.93e+01
43	46	69.7	1409	2	S41028	hypothetical protein	4.93e+01
44	45	68.2	527	2	A49865	protein-tyrosine kina	7.47e+01
45	45	68.2	544	2	S41093	triacylglycerol lipas	7.47e+01

ALIGNMENTS

```
RESULT     1
ENTRY            SNFF5K     #type complete
TITLE            multicatalytic endopeptidase complex (EC 3.4.99.46) 35K chain
                   - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES  19S cylinder particle 35K chain; multicatalytic proteinase
                   35K chain; prosome 35K chain; proteasome 35K chain
ORGANISM         #formal_name Drosophila melanogaster
DATE             30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
                   05-Sep-1997
ACCESSIONS       S23450; S05507; A38761
REFERENCE        S23450
   #authors      Frentzel, S.; Troxell, M.; Haass, C.; Pesold-Hurt, B.;
                   Glaetzer, K.H.; Kloetzel, P.M.
   #journal      Eur. J. Biochem. (1992) 205:1043-1051
   #title        Molecular characterization of the genomic regions of the
                   Drosophila alpha-type subunit proteasome genes PROS-Dm28.1
                   and PROS-Dm35.
   #cross-references MUID:92249308
   #accession    S23450
      ##molecule_type DNA
      ##residues      1-279 ##label FREN
      ##cross-references EMBL:X62285; NID:g8387; PID:g8388
      ##experimental_source strain Canton S
REFERENCE        S05507
   #authors      Haass, C.; Pesold-Hurt, B.; Multhaup, G.; Beyreuther, K.;
                   Kloetzel, P.M.
   #journal      EMBO J. (1989) 8:2373-2379
   #title        The PROS-35 gene encodes the 35 kd protein subunit of
                   Drosophila melanogaster proteasome.
   #cross-references MUID:90005444
   #accession    S05507
      ##molecule_type mRNA
      ##residues      1-279 ##label HAA
      ##cross-references EMBL:X15497; NID:g8381; PID:g8382
   #accession    A38761
      ##molecule_type protein
      ##residues      4-18;194-206 ##label HAA2
GENETICS
   #gene          PROS-35
      ##cross-references FlyBase:FBgn0003151
   #map_position 89F-90A
   #introns       1/3; 211/3
CLASSIFICATION   #superfamily multicatalytic endopeptidase complex chain C9
KEYWORDS         hydrolase; phosphoprotein; proteinase
SUMMARY          #length 279  #molecular-weight 31058  #checksum 365

  Query Match            86.4%;  Score 57;  DB 1;  Length 279;
```

```
  Best Local Similarity  75.0%;  Pred. No. 3.45e-01;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    121 QRYDRRPY 128
          |||:| ||
Qy      1 QRYNRAPY 8


RESULT    2
ENTRY             SNRTC2     #type complete
TITLE             multicatalytic endopeptidase complex (EC 3.4.99.46) chain C2
                    - rat
ALTERNATE_NAMES   multicatalytic proteinase component C2; proteasome chain C2
ORGANISM          #formal_name Rattus norvegicus #common_name Norway rat
DATE              31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
                    17-Oct-1997
ACCESSIONS        A32968; A38799; S09741
REFERENCE         A32968
   #authors       Fujiwara, T.; Tanaka, K.; Kumatori, A.; Shin, S.; Yoshimura,
                    T.; Ichihara, A.; Tokunaga, F.; Aruga, R.; Iwanaga, S.;
                    Kakizuka, A.; Nakanishi, S.
   #journal       Biochemistry (1989) 28:7332-7340
   #title         Molecular cloning of cDNA for proteasomes (multicatalytic
                    proteinase complexes) from rat liver: primary structure of
                    the largest component (C2).
   #cross-references MUID:90057428
   #accession     A32968
      ##molecule_type mRNA
      ##residues       1-263 ##label FUJ1
      ##cross-references EMBL:M29859; NID:g206381; PID:g206382
   #accession     A38799
      ##molecule_type protein
      ##residues       2-25;42-58;63-74,'X',76-79,'X',81;116-135;190-203;
                        218-226,'XX',229,'X',231;244-246,'X',248-262 ##label
                        FUJ2
REFERENCE         S09741
   #authors       Tokunaga, F.; Aruga, R.; Iwanaga, S.; Tanaka, K.; Ichihara,
                    A.; Takao, T.; Shimonishi, Y.
   #journal       FEBS Lett. (1990) 263:373-375
   #title         The NH2-terminal residues of rat liver proteasome
                    (multicatalytic proteinase complex) subunits, C2, C3 and
                    C8, are N-alpha-acetylated.
   #cross-references MUID:90243011
   #accession     S09741
      ##molecule_type protein
      ##residues       1-30 ##label TOK
CLASSIFICATION    #superfamily multicatalytic endopeptidase complex chain C9
KEYWORDS          acetylated amino end; hydrolase; proteinase
FEATURE
   1-263               #product multicatalytic endopeptidase complex chain C2
                         #status experimental #label MAT\
   1                   #modified_site acetylated amino end (Met) #status
                         experimental
SUMMARY           #length 263  #molecular-weight 29517  #checksum 9238

  Query Match           81.8%;  Score 54;  DB 1;  Length 263;
  Best Local Similarity  75.0%;  Pred. No. 1.43e+00;
  Matches     6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    121 QRYGRRPY 128
          ||| | ||
Qy      1 QRYNRAPY 8


RESULT    3
ENTRY             JC1445     #type complete
TITLE             multicatalytic endopeptidase complex (EC 3.4.99.46) chain C2,
                    long splice form - human
ALTERNATE_NAMES   macropain nu chain; multicatalytic endopeptidase complex HC2
                    chain; multicatalytic endopeptidase complex nu chain;
                    multicatalytic proteinase chain C2; prosome 30-33K chain;
                    proteasome alpha 1 subunit; proteasome chain C2; proteasome
                    nu chain
CONTAINS          multicatalytic endopeptidase complex chain C2, short splice
                    form
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                    08-Sep-1997
ACCESSIONS        JC1445; S15897; S17520; S25410; PC2321
REFERENCE         JC1445
   #authors       Pereira, I.S.; Bey, F.; Coux, O.; Scherrer, K.
   #journal       Gene (1992) 120:235-242
   #title         Two mRNAs exist for the Hs PROS-30 gene encoding a component
                    of human prosomes.
   #accession     JC1445
      ##molecule_type mRNA
      ##residues       1-269 ##label PER
      ##cross-references GB:M64992
REFERENCE         S15897
   #authors       Tamura, T.; Lee, D.H.; Osaka, F.; Fujiwara, T.; Shin, S.;
                    Chung, C.H.; Tanaka, K.; Ichihara, A.
   #journal       Biochim. Biophys. Acta (1991) 1089:95-102
   #title         Molecular cloning and sequence analysis of cDNAs for five
                    major subunits of human proteasomes (multi-catalytic
                    proteinase complexes).
   #cross-references MUID:91223105
   #accession     S15897
      ##molecule_type mRNA
      ##residues       'M',8-269 ##label TAM
      ##cross-references EMBL:D00759; NID:g220021; PID:d1001114; PID:g220022
REFERENCE         S17520
   #authors       DeMartino, G.N.; Orth, K.; McCullough, M.L.; Lee, L.W.; Munn,
                    T.Z.; Moomaw, C.R.; Dawson, P.A.; Slaughter, C.A.
   #journal       Biochim. Biophys. Acta (1991) 1079:29-38
   #title         The primary structures of four subunits of the human,
                    high-molecular-weight proteinase, macropain (proteasome),
                    are distinct but homologous.
   #cross-references MUID:91363412
   #accession     S17520
      ##status        not compared with conceptual translation
      ##molecule_type mRNA
      ##residues       'M',8-269 ##label DEM
      ##cross-references GB:X61969; NID:g296737; PID:g296738
   #accession     S25410
      ##molecule_type protein
      ##residues       10-40;46-61;68-75;89-95;103-128;132-148;164-168;176-195;
                          203-223;225-267 ##label DE2
REFERENCE         PC2315
   #authors       Kristensen, P.; Johnsen, A.H.; Uerkvitz, W.; Tanaka, K.;
                    Hendil, K.B.
   #journal       Biochem. Biophys. Res. Commun. (1994) 205:1785-1789
   #title         Human proteasome subunits from 2-dimensional gels identified
                    by partial sequencing.
   #cross-references MUID:95110324
   #accession     PC2321
      ##molecule_type protein
      ##residues       69-88 ##label KRI
      ##experimental_source placenta
COMMENT      The proteasome consists of subunits of 21K-30K arranged in 4
               stacked rings.
GENETICS
   #gene          GDB:PSMA1; HsPROS-30
      ##cross-references GDB:134040
   #map_position 11q-11q
CLASSIFICATION    #superfamily multicatalytic endopeptidase complex chain C9
KEYWORDS          hydrolase; phosphoprotein; proteinase
FEATURE
   2-269               #product multicatalytic endopeptidase complex chain C2,
                         long splice form #status predicted #label MATL\
   8-269               #product multicatalytic endopeptidase complex chain C2,
                         short splice form #status predicted #label MATS\
   230                 #binding_site phosphate (Tyr) (covalent) #status
                         predicted
SUMMARY           #length 269  #molecular-weight 30239  #checksum 5946

  Query Match           81.8%;  Score 54;  DB 2;  Length 269;
```

```
  Best Local Similarity  75.0%;  Pred. No. 1.43e+00;
  Matches     6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    127 QRYGRRPY 134
          ||| | ||
Qy      1 QRYNRAPY 8


RESULT    4
ENTRY             S77032     #type complete
TITLE             ABC-type transport protein s110778 - Synechocystis sp.
                    (strain PCC 6803)
ALTERNATE_NAMES   protein s110778
ORGANISM          #formal_name Synechocystis sp.
   #variety       PCC 6803
DATE              25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                    21-Aug-1998
ACCESSIONS        S77032
REFERENCE         S74322
   #authors       Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                    Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                    Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                    Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                    S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                    Yasuda, M.; Tabata, S.
   #journal       DNA Res. (1996) 3:109-136
   #title         Sequence analysis of the genome of the unicellular
                    cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                    determination of the entire genome and assignment of
                    potential protein-coding regions.
   #cross-references MUID:97061201
   #accession     S77032
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-790 ##label KAN
      ##cross-references EMBL:D64005; GB:AB001339; NID:g1001779; PID:d1011375;
                          PID:g1006575
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, June 1996
CLASSIFICATION    #superfamily ATP-binding cassette homology
KEYWORDS          P-loop; transport protein
FEATURE
   244-434             #domain ATP-binding cassette homology #label ABC\
   261-268             #region nucleotide-binding motif A (P-loop)
SUMMARY           #length 790  #molecular-weight 87656  #checksum 7314

  Query Match           80.3%;  Score 53;  DB 2;  Length 790;
  Best Local Similarity  62.5%;  Pred. No. 2.27e+00;
  Matches     5;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    475 QRYSRSSY 482
          |||:|::|
Qy      1 QRYNRAPY 8


RESULT    5
ENTRY             B71854     #type complete
TITLE             type II DNA modification enzyme (methyltransferase) -
                    Helicobacter pylori (strain J99)
ORGANISM          #formal_name Helicobacter pylori
   #variety       strain J99
DATE              12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                    12-Feb-1999
ACCESSIONS        B71854
REFERENCE         A71800
   #authors       Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                    Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                    B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                    Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                    Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                    G.F.; Trust, T.J.
   #journal       Nature (1999) 397:176-180
   #title         Genomic sequence comparison of two unrelated isolates of the
                    human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession     B71854
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-321 ##label ARN
      ##cross-references GB:AE001533; GB:AE001439; NID:g4155636; PID:g4155654
      ##experimental_source strain J99
GENETICS
   #gene          jhp1050
SUMMARY           #length 321  #molecular-weight 36919  #checksum 5430

  Query Match           78.8%;  Score 52;  DB 2;  Length 321;
  Best Local Similarity  75.0%;  Pred. No. 3.58e+00;
  Matches     6;  Conservative    0;  Mismatches   2;  Indels   0;  Gaps   0;

Db    204 QRYLRNPY 211
          ||| | ||
Qy      1 QRYNRAPY 8


RESULT    6
ENTRY             S28407     #type complete
TITLE             guanine nucleotide-exchange activator CDC25 homolog - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
                    10-Sep-1997
ACCESSIONS        S28407; S22693; B46199; S20730
REFERENCE         S28407
   #authors       Cen, H.; Papageorge, A.G.; Zippel, R.; Lowy, D.R.; Zhang, K.
   #journal       EMBO J. (1992) 11:4007-4015
   #title         Isolation of multiple mouse cDNAs with coding homology to
                    Saccharomyces cerevisiae CDC25: identification of a region
                    related to Bcr, Vav, Dbl and CDC24.
   #cross-references MUID:93010996
   #accession     S28407
      ##status         not compared with conceptual translation
      ##molecule_type mRNA
      ##residues       1-1260 ##label CEN
REFERENCE         S22693
   #authors       Martegani, E.; Vanoni, M.; Zippel, R.; Coccetti, P.;
                    Brambilla, R.; Ferrari, C.; Sturani, E.; Alberghina, L.
   #journal       EMBO J. (1992) 11:2151-2157
   #title         Cloning by functional complementation of a mouse cDNA
                    encoding a homologue of CDC25, a Saccharomyces cerevisiae
                    RAS activator.
   #cross-references MUID:92289680
   #accession     S22693
      ##molecule_type mRNA
      ##residues       789-1260 ##label MAR
      ##cross-references EMBL:X59868; NID:g50357; PID:g50358
REFERENCE         A46199
   #authors       Wei, W.; Mosteller, R.D.; Sanyal, P.; Gonzales, E.; McKinney,
                    D.; Dasgupta, C.; Li, P.; Liu, B.X.; Broek, D.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1992) 89:7100-7104
   #title         Identification of a mammalian gene structurally and
                    functionally related to the CDC25 gene of Saccharomyces
                    cerevisiae.
   #cross-references MUID:92357779
   #accession     B46199
      ##status         preliminary
      ##molecule_type nucleic acid
      ##residues       1029-1030,'D',1032-1224 ##label WEI
      ##experimental_source fetus
      ##note           sequence extracted from NCBI backbone (NCBIN:111101,
                        NCBIP:111102)
CLASSIFICATION    #superfamily CDC25-type guanine nucleotide exchange activator
                    homology; CDC24 homology; pleckstrin repeat homology
FEATURE
   242-428             #domain CDC24 homology #label CD24\
   1021-1257           #domain CDC25-type guanine nucleotide exchange activator
                         homology #label SOS
SUMMARY           #length 1260  #molecular-weight 143900  #checksum 9725
```

```
  Query Match           77.3%;  Score 51;  DB 2;  Length 1260;
  Best Local Similarity  50.0%;  Pred. No. 5.63e+00;
  Matches     4;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db   1058 EKYERTPY 1065
          ::|:|:||
Qy      1 QRYNRAPY 8


RESULT    7
ENTRY             G02630     #type complete
TITLE             FcalphaRb - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
                    17-Jul-1998
ACCESSIONS        G02630
REFERENCE         H01508
   #authors       van Dijk, T.B.; Morton, H.C.; Caldenhoven, E.; Bracke, M.;
                    Raaijmakers, J.A.M.; Lammers, J.
   #submission    submitted to the EMBL Data Library, April 1996
   #accession     G02630
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-239 ##label VAN
      ##cross-references EMBL:U56236; NID:g1326228; PID:g1326229
SUMMARY           #length 239  #molecular-weight 26996  #checksum 5338

  Query Match           75.8%;  Score 50;  DB 2;  Length 239;
  Best Local Similarity  71.4%;  Pred. No. 8.80e+00;
  Matches     5;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    196 WYNRSPY 202
          :|||:||
Qy      2 RYNRAPY 8


RESULT    8
ENTRY             JH0332     #type complete
TITLE             IgA (Fc) receptor, myeloid cell (CD89) precursor - human
ALTERNATE_NAMES   myeloid glycoprotein CD89
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
                    10-Sep-1997
ACCESSIONS        JH0332; I37224; S14405
REFERENCE         JH0332
   #authors       Maliszewski, C.R.; March, C.J.; Schoenborn, M.A.; Gimpel, S.;
                    Shen, L.
   #journal       J. Exp. Med. (1990) 172:1665-1672
   #title         Expression cloning of a human Fc receptor for IgA.
   #cross-references MUID:91079769
   #accession     JH0332
      ##molecule_type mRNA
      ##residues       1-287 ##label MAL
      ##cross-references GB:X54150; NID:g31329; PID:g31330
      ##experimental_source myeloid cell liver V937
REFERENCE         I37224
   #authors       de Wit, T.P.; Morton, H.C.; Capel, P.J.; van de Winkel, J.G.
   #journal       J. Immunol. (1995) 155:1203-1209
   #title         Structure of the gene for the human myeloid IgA Fc receptor
                    (CD89).
   #cross-references MUID:95363085
   #accession     I37224
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-287 ##label RES
      ##cross-references EMBL:X87767; NID:g963041; PID:g1054737
GENETICS
   #gene          GDB:FCAR; CD89
      ##cross-references GDB:127543; OMIM:147045
   #map_position 19q13.2-19q13.4
   #introns       12/1; 24/1; 121/1; 217/1
KEYWORDS          glycoprotein; immunoglobulin receptor; transmembrane protein
FEATURE
   1-21               #domain signal sequence #status predicted #label SIG\
   22-287             #product IgA receptor Fc alpha #status predicted #label
                        MFC\
   228-246            #domain transmembrane #status predicted #label TRA\
   65,79,141,177,186  #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY           #length 287  #molecular-weight 32265  #checksum 5963

  Query Match           75.8%;  Score 50;  DB 2;  Length 287;
  Best Local Similarity  71.4%;  Pred. No. 8.80e+00;
  Matches     5;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    196 WYNRSPY 202
          :|||:||
Qy      2 RYNRAPY 8


RESULT    9
ENTRY             I39956     #type complete
TITLE             neutral proteinase (EC 3.4.24.-) - Bacillus amyloliquefaciens
ORGANISM          #formal_name Bacillus amyloliquefaciens
DATE              19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change
                    18-Mar-1997
ACCESSIONS        I39956
REFERENCE         I39956
   #authors       Shimada, H.; Honjo, M.; Mita, I.; Nakayama, A.; Akaoka, A.;
                    Manabe, K.; Furutani, Y.
   #journal       J. Biotechnol. (1985) 2:75-85
   #title         The nucleotide sequence and some properties of the neutral
                    protease gene of Bacillus amyloliquefaciens.
   #accession     I39956
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-521 ##label RES
      ##cross-references GB:M36723; NID:g143352; PID:g143353
GENETICS
   #start_codon   GTG
CLASSIFICATION    #superfamily thermolysin
KEYWORDS          hydrolase; metalloproteinase
SUMMARY           #length 521  #molecular-weight 56725  #checksum 6816

  Query Match           75.8%;  Score 50;  DB 2;  Length 521;
  Best Local Similarity  62.5%;  Pred. No. 8.80e+00;
  Matches     5;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    308 QKYNRNSY 315
          |:||| :|
Qy      1 QRYNRAPY 8


RESULT   10
ENTRY             S52662     #type complete
TITLE             S-adenosylmethionine decarboxylase (SAMDC) - potato
ORGANISM          #formal_name Solanum tuberosum #common_name potato
DATE              14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change
                    21-Jul-1995
ACCESSIONS        S52662
REFERENCE         S52662
   #authors       Arif, S.A.M.; Taylor, M.A.; George, L.A.; Butler, A.R.;
                    Burch, L.R.; Davies, H.V.; Stark, M.J.R.; Kumar, A.
   #journal       Plant Mol. Biol. (1994) 26:327-338
   #title         Characterisation of the S-adenosylmethionine decarboxylase
                    (SAMDC) gene of potato.
   #accession     S52662
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-360 ##label ARI
SUMMARY           #length 360  #molecular-weight 39724  #checksum 8045

  Query Match           74.2%;  Score 49;  DB 2;  Length 360;
  Best Local Similarity  50.0%;  Pred. No. 1.37e+01;
  Matches     4;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db    332 QKFTRTPY 339
          |:: |:||
Qy      1 QRYNRAPY 8


RESULT   11
ENTRY            S28047     #type complete
TITLE            TUB13 protein - potato
ORGANISM         #formal_name Solanum tuberosum #common_name potato
DATE             17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
                   09-Sep-1997
ACCESSIONS       S28047
REFERENCE        S28046
   #authors      Taylor, M.A.; Arif, S.A.M.; Kumar, A.; Davies, H.V.; Scobie,
                   L.A.; Pearce, S.R.; Flavell, A.J.
   #journal      Plant Mol. Biol. (1992) 20:641-651
   #title        Expression and sequence analysis of cDNAs induced during the
                   early stages of tuberisation in different organs of the
                   potato plant (Solanum tuberosum L.).
   #accession    S28047
      ##molecule_type mRNA
      ##residues      1-360 ##label TAY
      ##cross-references EMBL:Z11680; NID:g21484; PID:g21485
GENETICS
   #gene         TUB13
SUMMARY          #length 360  #molecular-weight 39726  #checksum 7941

  Query Match           74.2%;  Score 49;  DB 2;  Length 360;
  Best Local Similarity  50.0%;  Pred. No. 1.37e+01;
  Matches     4;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    332 QKFTRTPY 339
          |:: |:||
Qy      1 QRYNRAPY 8


RESULT   12
ENTRY            S49009     #type complete
TITLE            fork head protein 2 - African clawed frog
ORGANISM         #formal_name Xenopus laevis #common_name African clawed frog
DATE             07-May-1995 #sequence_revision 21-Jul-1995 #text_change
                   17-Mar-1999
ACCESSIONS       S49009
REFERENCE        S49008
   #authors      Lef, J.; Clement, J.H.; Oschwald, R.; Koester, M.; Knoechel,
                   W.
   #journal      Mech. Dev. (1994) 45:117-126
   #title        Spatial and temporal transcription patterns of the forkhead
                   related XFD-2/XFD-2' genes in Xenopus laevis embryos.
   #cross-references MUID:94257528
   #accession    S49009
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-367 ##label LEF
      ##cross-references EMBL:X74316; NID:g511161; PID:g511162
CLASSIFICATION   #superfamily unassigned fork head proteins; fork head
                   DNA-binding domain homology
FEATURE
   128-219           #domain fork head DNA-binding domain homology #label FHD
SUMMARY          #length 367  #molecular-weight 40971  #checksum 6845

  Query Match           72.7%;  Score 48;  DB 2;  Length 367;
  Best Local Similarity  83.3%;  Pred. No. 2.11e+01;
  Matches     5;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    337 YNRSPY 342
          |||:||
Qy      3 YNRAPY 8


RESULT   13
ENTRY            S49008     #type complete
TITLE            fork head protein - African clawed frog
ORGANISM         #formal_name Xenopus laevis #common_name African clawed frog
DATE             07-May-1995 #sequence_revision 21-Jul-1995 #text_change
                   17-Mar-1999
ACCESSIONS       S49008; B56556
REFERENCE        S49008
   #authors      Lef, J.; Clement, J.H.; Oschwald, R.; Koester, M.; Knoechel,
                   W.
   #journal      Mech. Dev. (1994) 45:117-126
   #title        Spatial and temporal transcription patterns of the forkhead
                   related XFD-2/XFD-2' genes in Xenopus laevis embryos.
   #cross-references MUID:94257528
   #accession    S49008
      ##status       preliminary
      ##molecule_type mRNA
      ##residues      1-370 ##label LEF
      ##cross-references EMBL:X74315; NID:g511159; PID:g511160
REFERENCE        A56556
   #authors      Knochel, S.; Lef, J.; Clement, J.; Klocke, B.; Hille, S.;
                   Koster, M.; Knochel, W.
   #journal      Mech. Dev. (1992) 38:157-165
   #title        Activin A induced expression of a fork head related gene in
                   posterior chordamesoderm (notochord) of Xenopus laevis
                   embryos.
   #cross-references MUID:93041288
   #accession    B56556
      ##status       preliminary; not compared with conceptual translation
      ##molecule_type mRNA
      ##residues      118-228 ##label KNO
      ##experimental_source gastrula
      ##note         sequence extracted from NCBI backbone (NCBIP:118178)
CLASSIFICATION   #superfamily unassigned fork head proteins; fork head
                   DNA-binding domain homology
FEATURE
   127-218           #domain fork head DNA-binding domain homology #label FHD
SUMMARY          #length 370  #molecular-weight 41388  #checksum 6542

  Query Match           72.7%;  Score 48;  DB 2;  Length 370;
  Best Local Similarity  83.3%;  Pred. No. 2.11e+01;
  Matches     5;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    335 YNRSPY 340
          |||:||
Qy      3 YNRAPY 8


RESULT   14
ENTRY            S71800     #type fragment
TITLE            transcription factor FAST-1 - African clawed frog (fragment)
ALTERNATE_NAMES  forkhead activin signal transducer 1
ORGANISM         #formal_name Xenopus laevis #common_name African clawed frog
DATE             06-Dec-1996 #sequence_revision 25-Apr-1997 #text_change
                   05-Dec-1997
ACCESSIONS       S71800
REFERENCE        S71800
   #authors      Chen, X.; Rubock, M.J.; Whitman, M.
   #journal      Nature (1996) 383:691-696
   #title        A transcriptional partner for MAD proteins in TGF-beta
                   signalling.
   #cross-references MUID:97032727
   #accession    S71800
      ##status       nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-534 ##label CHE
      ##cross-references EMBL:U70980
CLASSIFICATION   #superfamily unassigned fork head proteins; fork head
                   DNA-binding domain homology
KEYWORDS         DNA binding; transcription factor
FEATURE
   117-209           #domain fork head DNA-binding domain homology #label FHD
SUMMARY          #length 534  #checksum 6119

  Query Match           72.7%;  Score 48;  DB 2;  Length 534;
```

```
  Best Local Similarity  50.0%;  Pred. No. 2.11e+01;
  Matches     4;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    113 HRYYKPPY 120
          :|| ::||
Qy      1 QRYNRAPY 8


RESULT   15
ENTRY             S74487     #type complete
TITLE             hypothetical protein sll1060 - Synechocystis sp. (strain PCC
                    6803)
ORGANISM          #formal_name Synechocystis sp.
   #variety       PCC 6803
DATE              25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                    21-Aug-1998
ACCESSIONS        S74487
REFERENCE         S74322
   #authors       Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                    Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                    Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                    Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                    S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                    Yasuda, M.; Tabata, S.
   #journal       DNA Res. (1996) 3:109-136
   #title         Sequence analysis of the genome of the unicellular
                    cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                    determination of the entire genome and assignment of
                    potential protein-coding regions.
   #cross-references MUID:97061201
   #accession     S74487
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-1032 ##label KAN
      ##cross-references EMBL:D90899; GB:AB001339; NID:g1651650; PID:d1017372;
                          PID:g1651711
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, June 1996
GENETICS
   #start_codon   GTG
SUMMARY           #length 1032  #molecular-weight 117162  #checksum 9339

  Query Match            72.7%;  Score 48;  DB 2;  Length 1032;
  Best Local Similarity  57.1%;  Pred. No. 2.11e+01;
  Matches     4;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    708 RFDRDPY 714
          |::| ||
Qy      2 RYNRAPY 8


Search completed: Thu Sep  2 11:17:44 1999
Job time : 14 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 11:16:54 1999; MasPar time 3.38 Seconds
56.561 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-3
Description: (1-9) from US08599226.pep
Perfect Score: 66
Sequence: 1 QRYNRAPYX 9

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.526; Variance 49.437; scale 0.334

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	66	100.0	9	27	W27585	Anti-TNF-alpha antibo	1.22e+00
2	66	100.0	9	27	W27562	Anti-TNF-alpha antibo	1.22e+00
3	66	100.0	107	27	W27568	Anti-TNF-alpha antibo	1.22e+00
4	61	92.4	9	27	W27579	Anti-TNF-alpha antibo	4.64e+00
5	61	92.4	9	27	W27575	Anti-TNF-alpha antibo	4.64e+00
6	61	92.4	9	27	W27571	Anti-TNF-alpha antibo	4.64e+00
7	58	87.9	9	27	W27584	Anti-TNF-alpha antibo	1.02e+01
8	56	84.8	9	27	W27572	Anti-TNF-alpha antibo	1.71e+01
9	54	81.8	263	4	R22957	Human proteasome comp	2.87e+01
10	54	81.8	269	1	R22666	Protein used to raise	2.87e+01
11	51	77.3	9	27	W27570	Anti-TNF-alpha antibo	6.14e+01
12	51	77.3	9	27	W27576	Anti-TNF-alpha antibo	6.14e+01
13	51	77.3	9	27	W27577	Anti-TNF-alpha antibo	6.14e+01
14	51	77.3	9	27	W27574	Anti-TNF-alpha antibo	6.14e+01
15	51	77.3	170	8	R42371	Prod. of ORF5 of plas	6.14e+01
16	50	75.8	287	11	R59920	Human Fc-alpha-R.	7.89e+01
17	50	75.8	287	7	R34030	Fc-alpha-R.	7.89e+01
18	49	74.2	360	13	R75006	Tomato S-adenosyl-met	1.01e+02
19	47	71.2	9	27	W27582	Anti-TNF-alpha antibo	1.66e+02
20	47	71.2	9	27	W27573	Anti-TNF-alpha antibo	1.66e+02
21	47	71.2	419	28	W26744	S. carnosus nitrate r	1.66e+02
22	47	71.2	447	3	R12388	smg p25A GDP Dissocia	1.66e+02
23	46	69.7	9	27	W27578	Anti-TNF-alpha antibo	2.12e+02
24	46	69.7	67	31	W28077	Amino acid sequence o	2.12e+02
25	45	68.2	246	13	R71131	Tyrosine kinase domai	2.70e+02
26	45	68.2	466	13	R71132	N-terminal truncated	2.70e+02
27	45	68.2	507	13	R71133	Cytoplsmic tyrosine k	2.70e+02
28	45	68.2	507	15	R84181	Megakaryocyte kinase	2.70e+02
29	45	68.2	520	1	P94617	Neutral protease enco	2.70e+02
30	45	68.2	521	3	P51009	Sequence of neutral p	2.70e+02
31	45	68.2	528	34	W64454	Human matK protein.	2.70e+02
32	45	68.2	557	2	R06370	Protein with lipase a	2.70e+02
33	45	68.2	563	9	R47577	Lipase of Geotrichum	2.70e+02
34	45	68.2	563	2	R10330	Gene product with lip	2.70e+02
35	44	66.7	9	27	W27580	Anti-TNF-alpha antibo	3.43e+02
36	44	66.7	107	39	W86127	Murine 708 Vl amino a	3.43e+02
37	44	66.7	107	39	W86129	Protein sequence of d	3.43e+02
38	44	66.7	287	19	W06702	Partial medium chain-	3.43e+02
39	44	66.7	400	10	R56281	Chitin-deacetylase.	3.43e+02
40	44	66.7	454	32	W58586	Human histamine H2 re	3.43e+02
41	44	66.7	517	20	W04558	Carnation ACC synthas	3.43e+02
42	44	66.7	529	36	W70501	Human G-protein adren	3.43e+02
43	44	66.7	2251	14	R71009	Human neuronal calciu	3.43e+02
44	44	66.7	2270	14	R71010	Human neuronal calciu	3.43e+02
45	44	66.7	2408	2	R24305	Translation of ORF 1	3.43e+02

ALIGNMENTS

RESULT 1
ID W27585 standard; peptide; 9 AA.
AC W27585;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 72; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 66;  DB 27;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches     8;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 qrynrapy 8
          ||||||||
Qy      1 QRYNRAPY 8


RESULT    2
ID   W27562 standard; peptide; 9 AA.
AC   W27562;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody light chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   light chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 9
FT                   /label= Thr, Ala
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
DR   N-PSDB; T88403.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 9; Page 64; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 66;  DB 27;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches     8;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 qrynrapy 8
          ||||||||
Qy      1 QRYNRAPY 8


RESULT    3
ID   W27568 standard; Protein; 107 AA.
AC   W27568;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody light chain variable region.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody;
KW   light chain; variable region; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
DR   N-PSDB; T88403.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 15; Page 75; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) light chain variable region.
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   107 AA;

  Query Match          100.0%;  Score 66;  DB 27;  Length 107;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches     8;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     89 qrynrapy 96
          ||||||||
Qy      1 QRYNRAPY 8


RESULT    4
ID   W27579 standard; peptide; 9 AA.
AC   W27579;
```

DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 70; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbance,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 92.4%; Score 61; DB 27; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.64e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 qkynrapy 8
|:||||||
Qy 1 QRYNRAPY 8

RESULT 5
ID W27575 standard; peptide; 9 AA.
AC W27575;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 69; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 92.4%; Score 61; DB 27; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.64e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 qkynrapy 8
|:||||||
Qy 1 QRYNRAPY 8

RESULT 6
ID W27571 standard; peptide; 9 AA.
AC W27571;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 68; 102pp; English.

CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 92.4%; Score 61; DB 27; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.64e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 qkynrapy 8
|:||||||
Qy 1 QRYNRAPY 8

RESULT 7
ID W27584 standard; peptide; 9 AA.
AC W27584;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 72; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 87.9%; Score 58; DB 27; Length 9;
Best Local Similarity 75.0%; Pred. No. 1.02e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 qkynrppy 8
|:|||:||
Qy 1 QRYNRAPY 8

RESULT 8
ID W27572 standard; peptide; 9 AA.
AC W27572;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 68; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 84.8%; Score 56; DB 27; Length 9;
Best Local Similarity 75.0%; Pred. No. 1.71e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Db      1 qkyqrapy 8
          |:| ||||
Qy      1 QRYNRAPY 8
```

RESULT 9
ID R22957 standard; Protein; 263 AA.
AC R22957;
DT 15-OCT-1992 (first entry)
DE Human proteasome component HC2.
KW HC2; HC5; HC8; HC9; Alzheimer's disease; cancer.
OS Homo sapiens.
PN J04077498-A.
PD 11-MAR-1992.
PF 20-JUL-1990; 193313.
PR 20-JUL-1990; JP-193313.
PA (SAKA) OTSUKA PHARM KK.
DR WPI; 92-136767/17.
DR N-NSDB; Q23894.
PT Human proteasome - has specified aminoacid sequence and gene base
PT sequence, used for e.g. investigation of Alzheimer's disease
PS Claim 1; Page 1; 24pp; Japanese.
CC Human proteasome HC2 and its gene can be used in the investigation,
CC diagnosis, and treatment of diseases associated with proteasome
CC abnormality such as cancers and Alzheimers disease.
CC Human proteasome is purified from the soluble liver fractions of
CC human liver cell HepG2 cell. A probe was prepared, and a cDNA
CC library constructed using plasmid Bluescript KS+. A clone was
CC isolated and its DNA sequenced to give Q23894.
CC See also Q23894-7, R22957-60.
SQ Sequence 263 AA;

Query Match 81.8%; Score 54; DB 4; Length 263;
Best Local Similarity 75.0%; Pred. No. 2.87e+01;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Db    121 qrygrrpy 128
          ||| | ||
Qy      1 QRYNRAPY 8
```

RESULT 10
ID R22666 standard; Protein; 269 AA.
AC R22666;
DT 06-NOV-1992 (first entry)
DE Protein used to raise anti-p33K antibodies.
KW HIV; human immunodeficiency virus; promosomal; cell surface; p27K.
OS Homo sapiens.
PN WO9207269-A.
PD 30-APR-1992.
PF 10-OCT-1991; E01945.
PR 11-OCT-1990; EP-402838.
PA (PROS-) PRO-SOMA SARL.
PI Bey F, Bureau JP, Scherrer K;
DR WPI; 92-167288/20.
DR N-PSDB; Q24128.
PT Method for diagnosing HIV infection - comprises using
PT immunochemical reagent contg. monoclonal antibodies against
PT prosomal cell surface proteins
PS Claim 13; Fig 9; 34pp; English.
CC The 269 amino acid sequence is used to raise antibodies against the
CC prosomal surface protein p33K. The anti-p33K antibodies raised against
CC this protein can detect prosomal surface antigens. HIV infected cells
CC express these proteins on their surface, thus the antibodies may be used
CC to detect HIV infection and to determine whether T4 cells have been
CC infected even before they are killed by HIV.
CC See also R22665.
SQ Sequence 269 AA;

Query Match 81.8%; Score 54; DB 1; Length 269;
Best Local Similarity 75.0%; Pred. No. 2.87e+01;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Db    127 qrygrrpy 134
          ||| | ||
Qy      1 QRYNRAPY 8
```

RESULT 11
ID W27570 standard; peptide; 9 AA.
AC W27570;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 67; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 77.3%; Score 51; DB 27; Length 9;
Best Local Similarity 75.0%; Pred. No. 6.14e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Db      1 qkynsapy 8
          |:|| |||
Qy      1 QRYNRAPY 8
```

RESULT 12
ID W27576 standard; peptide; 9 AA.
AC W27576;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;

KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 69; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 77.3%; Score 51; DB 27; Length 9;
Best Local Similarity 75.0%; Pred. No. 6.14e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 qkynsapy 8
|:|| |||
Qy 1 QRYNRAPY 8

RESULT 13
ID W27577 standard; peptide; 9 AA.
AC W27577;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 70; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 77.3%; Score 51; DB 27; Length 9;
Best Local Similarity 75.0%; Pred. No. 6.14e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 qkynsapy 8
|:|| |||
Qy 1 QRYNRAPY 8

RESULT 14
ID W27574 standard; peptide; 9 AA.
AC W27574;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody light chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW light chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 69; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) light chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or

CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 9 AA;

Query Match 77.3%; Score 51; DB 27; Length 9;
Best Local Similarity 75.0%; Pred. No. 6.14e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
      1 qkynsapy 8
        |:|| |||
Qy      1 QRYNRAPY 8
```

RESULT 15
ID R42371 standard; Protein; 170 AA.
AC R42371;
DT 19-APR-1994 (first entry)
DE Prod. of ORF5 of plasmid pRAP501.
KW Haemophilus somnus; immunogenic; haemolysin; LppB; LppC;
KW thromboembolic meningoencephalitis; septicaemia; arthritis;
KW pneumonia; haemin-binging protein.
OS Haemophilus somnus.
PN WO9321323-A.
PD 28-OCT-1993.
PF 05-APR-1993; CA0135.
PR 09-APR-1992; US-865050.
PR 04-JUN-1992; US-893424.
PR 04-JUN-1992; US-893426.
PR 29-MAR-1993; US-038287.
PR 29-MAR-1993; US-038288.
PR 29-MAR-1993; US-038719.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
PI Rioux C, Theisen M;
DR WPI; 93-351733/44.
DR N-PSDB; Q51080.
PT Haemophilus somnus immunogenic proteins used in vaccines -
PT selected from haemin-binding protein, haemolysin, LppB and LppC,
PT and corresp. DNA
PS Disclosure; Fig 2; 119pp; English.
CC A genomic cosmid library of Haemophilus somnus HS25 DNA was screened
CC for clones capable of binding bovine haemin and having haemolytic
CC activity. Positive clones were subcloned various times, resulting
CC in plasmid pRAP501, which binds haemin but is not haemolytic. The
CC clone was sequenced and was found to contain several open reading
CC frames, potentially encoding 8 proteins. The haemin binding protein
CC (encoded by the hmb gene) was encoded by ORF1. The protein can be
CC used in vaccines for preventing or treating H. somnus infections,
CC which cause thromboembolic meningo-encephalitis, septicaemia,
CC arthritis and pneumonia in vertebrates. The protein shown is
CC potentially encoded by ORF5, and has an unknown function.
CC See also R42370-86.
SQ Sequence 170 AA;

Query Match 77.3%; Score 51; DB 8; Length 170;
Best Local Similarity 75.0%; Pred. No. 6.14e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Db    107 qrykravy 114
          |||:|| |
```

```
Qy      1 QRYNRAPY 8
```

Search completed: Thu Sep 2 11:17:12 1999
Job time : 18 secs.

****

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

[redacted] on: Thu Sep 2 11:20:18 1999; MasPar time 3.47 Seconds
73.528 Million cell updates/sec
[redacted]ular output not generated.

Title: >US-08-599-226-4
Description: (1-12) from US08599226.pep
Perfect Score: 66
Sequence: 1 VSYLSTASSLDX 12

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.837; Variance 49.496; scale 0.340

[redacted] Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	66	100.0	12	27	W27594	Anti-TNF-alpha antibo	1.22e+00
2	66	100.0	12	27	W27563	Anti-TNF-alpha antibo	1.22e+00
3	66	100.0	121	27	W27569	Anti-TNF-alpha antibo	1.22e+00
4	58	87.9	12	27	W27593	Anti-TNF-alpha antibo	1.04e+01
5	57	86.4	12	27	W27586	Anti-TNF-alpha antibo	1.35e+01
6	57	86.4	12	27	W27587	Anti-TNF-alpha antibo	1.35e+01
7	57	86.4	12	27	W27589	Anti-TNF-alpha antibo	1.35e+01
8	57	86.4	12	27	W27588	Anti-TNF-alpha antibo	1.35e+01
9	52	78.8	441	32	W28236	Amino acid sequence o	4.92e+01
10	51	77.3	12	27	W27592	Anti-TNF-alpha antibo	6.34e+01
11	51	77.3	12	27	W27591	Anti-TNF-alpha antibo	6.34e+01
12	49	74.2	12	27	W27590	Anti-TNF-alpha antibo	1.05e+02
13	48	72.7	417	26	W23067	Canine IgE heavy chai	1.35e+02
14	48	72.7	481	30	W40054	P300/CBP-associated t	1.35e+02
15	48	72.7	832	30	W40052	Human P300/CBP-associ	1.35e+02
16	48	72.7	2799	39	W81867	Human tumour suppress	1.35e+02
17	47	71.2	9	12	R61854	MBP peptide 14, poten	1.73e+02
18	47	71.2	10	12	R61948	MBP peptide 14, poten	1.73e+02
19	47	71.2	11	12	R61957	MBP peptide 14, poten	1.73e+02
20	47	71.2	13	16	R85136	Human MBP residues 13	1.73e+02
21	47	71.2	19	19	R95335	MBP-1.1 (11-29).	1.73e+02
22	47	71.2	19	27	W37551	Human myelin basic pr	1.73e+02
23	47	71.2	20	39	W73609	Human myelin basic pr	1.73e+02
24	47	71.2	20	19	R95334	MBP-1 (11-30).	1.73e+02
25	47	71.2	20	27	W37550	Human myelin basic pr	1.73e+02
26	47	71.2	21	35	W78823	Myelin basic protein	1.73e+02
27	47	71.2	21	19	R95336	MBP-1.2 (11-31).	1.73e+02
28	47	71.2	127	9	R48596	Rat myelin basic prot	1.73e+02
29	47	71.2	142	35	W38731	Streptococcus pneumon	1.73e+02
30	47	71.2	167	9	R48595	Myelin basic protein.	1.73e+02
31	47	71.2	168	9	R48594	Rabbit myelin basic p	1.73e+02
32	47	71.2	168	1	R04717	Empirically determine	1.73e+02
33	47	71.2	170	6	R30736	Human MBP.	1.73e+02
34	47	71.2	170	19	R95406	Myelin oligodendrocyt	1.73e+02
35	47	71.2	170	7	R35440	Human basic myelin pr	1.73e+02
36	47	71.2	170	9	R48592	Human myelin basic pr	1.73e+02
37	47	71.2	171	18	R97627	Human myelin basic pr	1.73e+02
38	47	71.2	171	18	R99580	Human myelin basic pr	1.73e+02
39	47	71.2	197	19	W00399	Human myelin basic pr	1.73e+02
40	47	71.2	203	19	W06108	Foetal myelin basic p	1.73e+02
41	47	71.2	203	19	W06107	Foetal myelin basic p	1.73e+02
42	47	71.2	373	19	W06103	MP4 chimera (MBP21.5-	1.73e+02
43	47	71.2	375	19	W06104	PM4 chimera (delta PL	1.73e+02
44	47	71.2	385	19	W06102	MP3 chimera (MBP21.5-	1.73e+02
45	47	71.2	492	19	W06105	MMOGP4 chimera (MBP21	1.73e+02

ALIGNMENTS

RESULT 1
ID W27594 standard; peptide; 12 AA.
AC W27594;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Disclosure; Page 75; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          100.0%;  Score 66;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches   11;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 vsylstassld 11
          ||||||||||
Qy      1 VSYLSTASSLD 11


RESULT    2
ID   W27563 standard; peptide; 12 AA.
AC   W27563;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 12
FT                   /label= Tyr, Asn
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 9; Page 65; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          100.0%;  Score 66;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches   11;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 vsylstassld 11
          ||||||||||
Qy      1 VSYLSTASSLD 11


RESULT    3
ID   W27569 standard; Protein; 121 AA.
AC   W27569;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain variable region.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody;
KW   heavy chain; variable region; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
DR   N-PSDB; T88404.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 16; Page 76; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain variable region.
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   121 AA;

  Query Match          100.0%;  Score 66;  DB 27;  Length 121;
  Best Local Similarity 100.0%;  Pred. No. 1.22e+00;
  Matches   11;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     99 vsylstassld 109
          ||||||||||
Qy      1 VSYLSTASSLD 11


RESULT    4
ID   W27593 standard; peptide; 12 AA.
AC   W27593;
```

DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 75; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 87.9%; Score 58; DB 27; Length 12;
Best Local Similarity 90.0%; Pred. No. 1.04e+01;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 2 sylstassle 11
||||||||:
Qy 2 SYLSTASSLD 11

RESULT 5
ID W27586 standard; peptide; 12 AA.
AC W27586;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 72; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 86.4%; Score 57; DB 27; Length 12;
Best Local Similarity 90.0%; Pred. No. 1.35e+01;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 2 sylstsssld 11
|||||:||||
Qy 2 SYLSTASSLD 11

RESULT 6
ID W27587 standard; peptide; 12 AA.
AC W27587;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.

CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 86.4%; Score 57; DB 27; Length 12;
Best Local Similarity 90.0%; Pred. No. 1.35e+01;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 2 sylstsssld 11
|||||:||||
Qy 2 SYLSTASSLD 11

RESULT 7
ID W27589 standard; peptide; 12 AA.
AC W27589;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 86.4%; Score 57; DB 27; Length 12;
Best Local Similarity 90.0%; Pred. No. 1.35e+01;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 2 sylstsssld 11
|||||:||||
Qy 2 SYLSTASSLD 11

RESULT 8
ID W27588 standard; peptide; 12 AA.
AC W27588;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 86.4%; Score 57; DB 27; Length 12;
Best Local Similarity 90.0%; Pred. No. 1.35e+01;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Db      2 sylstsssld 11
          |||||:||||
Qy      2 SYLSTASSLD 11


RESULT    9
ID   W28236 standard; Protein; 441 AA.
AC   W28236;
DT   07-SEP-1998  (first entry)
DE   Amino acid sequence of a mercuric reductase.
KW   Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW   Staphylococcal gene; regulatory element; bacterial gene expression;
KW   vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW   toxic shock syndrome; mercuric reductase.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   Misc_difference 16
FT                   /note= "Residue not specified"
PN   WO9730070-A1.
PD   21-AUG-1997.
PF   19-FEB-1997; U02318.
PR   20-FEB-1996; US-011888.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR   WPI; 97-424969/39.
DR   N-PSDB; T84151.
PT   Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT   to isolate antimicrobial compounds, and in vaccines against S.
PT   aureus infection
PS   Claim 6; Pages 547-548; 989pp; English.
CC   The present sequence represents a Staphylococcus aureus protein, that,
CC   based on homology with another Staphylococcus aureus protein, is
CC   believed to be a mercuric reductase. The DNA sequence was isolated from
CC   a library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
CC   sequence can be used in the construction of ribozymes and antisense
CC   sequences to control the expression of Staphylococcal genes. The DNA
CC   sequence is also useful as a source of regulatory elements for the
CC   control of bacterial gene expression. The present protein may be used
CC   to produce vaccines to enable a host to produce specific antibodies
CC   with antibacterial action. These vaccines and antibodies would protect
CC   a host against invasion by S. aureus, and conditions relating to
CC   Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC   skin syndrome, and toxic shock syndrome.
SQ   Sequence   441 AA;

  Query Match           78.8%;  Score 52;  DB 32;  Length 441;
  Best Local Similarity  45.5%;  Pred. No. 4.92e+01;
  Matches    5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    203 vdyltstsale 213
          | ||:::|:|:
Qy      1 VSYLSTASSLD 11


RESULT   10
ID   W27592 standard; peptide; 12 AA.
AC   W27592;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           77.3%;  Score 51;  DB 27;  Length 12;
  Best Local Similarity  70.0%;  Pred. No. 6.34e+01;
  Matches    7;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 sflstsssle 11
          |:|||:|||:
Qy      2 SYLSTASSLD 11


RESULT   11
ID   W27591 standard; peptide; 12 AA.
AC   W27591;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
```

```
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           77.3%;  Score 51;  DB 27;  Length 12;
  Best Local Similarity  88.9%;  Pred. No. 6.34e+01;
  Matches    8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 sylstsssl 10
          |||||:|||
Qy      2 SYLSTASSL 10


RESULT   12
ID   W27590 standard; peptide; 12 AA.
AC   W27590;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           74.2%;  Score 49;  DB 27;  Length 12;
  Best Local Similarity  80.0%;  Pred. No. 1.05e+02;
  Matches    8;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db      2 sylstsfsld 11
          |||||: |||
Qy      2 SYLSTASSLD 11


RESULT   13
ID   W23067 standard; Protein; 417 AA.
AC   W23067;
DT   19-FEB-1998  (first entry)
DE   Canine IgE heavy chain constant region (exon 1-4 product).
KW   IgE; immunoglobulin; antibody; heavy chain constant region;
KW   allergy; hypersensitivity; therapy; dog; antisense;
KW   immunomodulation.
OS   Canis familiaris.
FH   Key             Location/Qualifiers
FT   Misc_difference 55
FT                   /note= "encoded by ACC"
FT   Misc_difference 56
FT                   /note= "encoded by TAC"
FT   Misc_difference 67
FT                   /note= "encoded by GCC"
FT   Misc_difference 83
FT                   /note= "encoded by NNT"
FT   Misc_difference 174
FT                   /note= "encoded by GGN"
FT   Misc_difference 175
FT                   /note= "encoded by NNG"
FT   Misc_difference 176
FT                   /note= "encoded by TGN"
FT   Misc_difference 203
FT                   /note= "encoded by TCC"
FT   Misc_difference 204
FT                   /note= "encoded by GAC"
PN   WO9730156-A2.
PD   21-AUG-1997.
PF   14-FEB-1997; U02322.
PR   14-FEB-1996; US-601197.
PA   (IDEX-) IDEXX LAB INC.
PI   Harris RA,  Mermer B,  Siefring AE;
DR   WPI; 97-425031/39.
DR   N-PSDB; T79278.
PT   Isolated canine IgE heavy chain constant region DNA - useful to
PT   develop products for treatment of canine allergies and for
PT   immunomodulation in dogs
PS   Disclosure; Page 35-39; 59pp; English.
CC   This polypeptide is encoded by exons 1-4 (see T79278) of canine
CC   IgE heavy chain constant region (epsilon) genomic DNA.  Another
CC   polypeptide, comprising the exon 5 and 6 product, is given in
CC   W23068.  Recombinant peptides encoded by exons 1-6 can be
CC   produced in eukaryotic or prokaryotic cells.  Such peptides,
CC   and antibodies raised against them, are used in methods to treat
CC   the manifestation of allergy in dogs, e.g. to treatment Type I
CC   immediate hypersensitivity, and for immunomodulation.
SQ   Sequence   417 AA;

  Query Match           72.7%;  Score 48;  DB 26;  Length 417;
  Best Local Similarity  70.0%;  Pred. No. 1.35e+02;
  Matches    7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    210 sylsppspld 219
          |||| :|:||
Qy      2 SYLSTASSLD 11
```

```
RESULT   14
ID   W40054 standard; Protein; 481 AA.
AC   W40054;
DT   20-JUL-1998  (first entry)
DE   P300/CBP-associated transcriptional cofactor P/CAF C-terminus.
KW   P/CAF; human; p300; CBP; transcription; cofactor;
KW   histone acetyltransferase; HIV; infection; cancer; therapy;
KW   muscle differentiation.
OS   Homo sapiens.
PN   WO9803652-A2.
PD   29-JAN-1998.
PF   23-JUL-1997; U12877.
PR   23-JUL-1996; US-022273.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Howard BH, Nakatani Y;
DR   WPI; 98-120777/11.
PT   New isolated p300/CBP-associated factor, P/CAF - used to develop
PT   products for modulating transcription, e.g. for treating HIV
PT   infection or cancers or for promoting muscle differentiation
PS   Claim 2; Page 72-73; 107pp; English.
CC   This polypeptide comprises the C-terminal region (amino acid
CC   residues 352-832) of a novel human p300/CBP associated cofactor,
CC   P/CAF (see W40052), that modulates transcription through binding
CC   to the cell transcription cofactors p300 (see W40055) and CBP
CC   (see W40060) and through acetylation of histones. This C-terminal
CC   region contains the histone acetyltransferase activity of the
CC   protein. The invention provides methods of screening for compounds
CC   that inhibit or stimulate the transcription modulating and histone
CC   acetyltransferase activity of P/CAF and p300/CBP. Inhibitors can
CC   be used e.g. to inhibit HIV TAT-mediated transcription in the
CC   treatment of HIV infection. Stimulators can be used e.g. to
CC   activate tumour suppressor p53 in the treatment of cancer or to
CC   activate the muscle differentiation factor MyoD to promote muscle
CC   differentiation. The products can also be used to inhibit the cell
CC   cycle progression inducing effect of an oncoprotein which binds
CC   p300/CBP in a subject. The C-terminal fragment is specific for
CC   P/CAF and can be used to identify and define P/CAF.
SQ   Sequence   481 AA;

  Query Match           72.7%;  Score 48;  DB 30;  Length 481;
  Best Local Similarity  63.6%;  Pred. No. 1.35e+02;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     43 isynstsssle 53
          :|| ||:|||:
Qy      1 VSYLSTASSLD 11


RESULT   15
ID   W40052 standard; Protein; 832 AA.
AC   W40052;
DT   20-JUL-1998  (first entry)
DE   Human P300/CBP-associated transcriptional cofactor P/CAF.
KW   P/CAF; human; p300; CBP; transcription; cofactor;
KW   histone acetyltransferase; HIV; infection; cancer; therapy;
KW   muscle differentiation.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Region          1..351
FT                   /note= "contains binding site for p300/CBP
FT                         (Claim 2)"
FT   Region          352..832
FT                   /note= "contains histone acetyltransferase
FT                         activity (Claim 8)"
PN   WO9803652-A2.
PD   29-JAN-1998.
PF   23-JUL-1997; U12877.
PR   23-JUL-1996; US-022273.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Howard BH, Nakatani Y;
DR   WPI; 98-120777/11.
PT   New isolated p300/CBP-associated factor, P/CAF - used to develop
PT   products for modulating transcription, e.g. for treating HIV
PT   infection or cancers or for promoting muscle differentiation
PS   Claim 1; Page 71-72; 107pp; English.
CC   This protein is a novel human p300/CBP associated cofactor,
CC   designated P/CAF, that modulates transcription through binding to
CC   the cell transcription cofactors p300 (see W40055) and CBP (see
CC   W40060) and through acetylation of histones. Its amino acid
CC   sequence was deduced from a clone (see V10090) obtained from a
CC   human foetal liver cDNA library. The N-terminal region (see
CC   W40056) of P/CAF, which contains the binding site for p300/CBP,
CC   and the C-terminal region (see W40054), which contains the histone
CC   acetyltransferase activity are also claimed. The invention provides
CC   methods of screening for compounds that inhibit or stimulate the
CC   transcription modulating and histone acetyltransferase activity of
CC   P/CAF and p300/CBP. Inhibitors can be used e.g. to inhibit HIV
CC   TAT-mediated transcription in the treatment of HIV infection.
CC   Stimulators can be used e.g. to activate tumour suppressor p53 in
CC   the treatment of cancer or to activate the muscle differentiation
CC   factor MyoD to promote muscle differentiation. The products can
CC   also be used to inhibit the cell cycle progression inducing effect
CC   of an oncoprotein which binds p300/CBP in a subject. The products
CC   can also be used for detection, screening assays and production of
CC   transgenic animals.
SQ   Sequence   832 AA;

  Query Match           72.7%;  Score 48;  DB 30;  Length 832;
  Best Local Similarity  63.6%;  Pred. No. 1.35e+02;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    394 isynstsssle 404
          :|| ||:|||:
Qy      1 VSYLSTASSLD 11


Search completed: Thu Sep  2 11:20:37 1999
Job time : 19 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:30:26 1999; MasPar time 3.46 Seconds
73.740 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-29
Description: (1-12) from US08599226.pep
Perfect Score: 75
Sequence: 1 ASYLSTSSSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.262; Variance 57.060; scale 0.303

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	75	100.0	12	27	W27588	Anti-TNF-alpha antibo	4.55e-01
2	72	96.0	12	27	W27593	Anti-TNF-alpha antibo	9.58e-01
3	69	92.0	12	27	W27592	Anti-TNF-alpha antibo	2.00e+00
4	69	92.0	12	27	W27591	Anti-TNF-alpha antibo	2.00e+00
5	69	92.0	121	27	W27569	Anti-TNF-alpha antibo	2.00e+00
6	67	89.3	12	27	W27590	Anti-TNF-alpha antibo	3.26e+00
7	63	84.0	12	27	W27586	Anti-TNF-alpha antibo	8.55e+00
8	63	84.0	12	27	W27589	Anti-TNF-alpha antibo	8.55e+00
9	63	84.0	12	27	W27587	Anti-TNF-alpha antibo	8.55e+00
10	57	76.0	12	27	W27563	Anti-TNF-alpha antibo	3.53e+01
11	57	76.0	12	27	W27594	Anti-TNF-alpha antibo	3.53e+01
12	54	72.0	159	37	W80222	Human succinate-ubiqu	7.08e+01
13	54	72.0	159	29	W55260	H. pylori ORF 06cp306	7.08e+01
14	54	72.0	177	29	W55472	H. pylori ORF 06cp306	7.08e+01
15	54	72.0	526	29	W55688	H. pylori ORF 09cp107	7.08e+01
16	54	72.0	875	28	W34987	Bankia gouldi endoglu	7.08e+01
17	50	66.7	166	32	W59345	Human ErbB2 domain 1	1.76e+02
18	50	66.7	191	32	W59354	Human ErbB2 protein f	1.76e+02
19	50	66.7	417	26	W23067	Canine IgE heavy chai	1.76e+02
20	50	66.7	782	23	W19764	Her2-GM-CSF immunosti	1.76e+02
21	50	66.7	1255	19	W01111	HER-2/neu protein.	1.76e+02
22	50	66.7	1433	8	R39568	Sequence of c-erbB-2	1.76e+02
23	49	65.3	37	22	W17907	Peptide CS4 from dena	2.20e+02
24	49	65.3	37	26	W24223	Peptide fragment from	2.20e+02
25	49	65.3	117	22	W17913	Peptide CS4 from dena	2.20e+02
26	49	65.3	384	20	W06968	M. tuberculosis ORF 2	2.20e+02
27	49	65.3	409	34	W71495	Helicobacter polypept	2.20e+02
28	49	65.3	624	2	R08222	Extracellular portion	2.20e+02
29	49	65.3	1026	11	R58906	Human protocadherin-4	2.20e+02
30	49	65.3	1026	17	R87146	Protocadherin clone 4	2.20e+02
31	49	65.3	1203	17	R87152	Alternatively spliced	2.20e+02
32	49	65.3	1203	11	R58911	Product of alternativ	2.20e+02
33	48	64.0	459	27	W36485	Mouse TUB Form I.	2.74e+02
34	48	64.0	460	27	W36488	Human TUB Form 6.	2.74e+02
35	48	64.0	505	24	W10728	Mouse tub gene produc	2.74e+02
36	48	64.0	505	27	W36486	Mouse TUB Form II.	2.74e+02
37	48	64.0	505	32	W54367	Mouse tub polypeptide	2.74e+02
38	48	64.0	506	24	W10729	Human tub gene produc	2.74e+02
39	48	64.0	506	27	W36497	Human TUB Form 4.	2.74e+02
40	48	64.0	506	32	W54368	Human tub polypeptide	2.74e+02
41	48	64.0	512	27	W36496	Human TUB Form 3.	2.74e+02
42	48	64.0	518	27	W36495	Human TUB Form 2.	2.74e+02
43	48	64.0	561	27	W36489	Human TUB Form 1.	2.74e+02
44	48	64.0	2799	39	W81867	Human tumour suppress	2.74e+02
45	47	62.7	290	18	R97837	Kaposi's sarcoma asso	3.42e+02

ALIGNMENTS

RESULT 1
ID W27588 standard; peptide; 12 AA.
AC W27588;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          100.0%;  Score 75;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 4.55e-01;
  Matches    12;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssldy 12
          ||||||||||||
Qy      1 ASYLSTSSSLDY 12


RESULT    2
ID   W27593 standard; peptide; 12 AA.
AC   W27593;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 75; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          96.0%;  Score 72;  DB 27;  Length 12;
  Best Local Similarity 83.3%;  Pred. No. 9.58e-01;
  Matches    10;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstassley 12
          ||||||:|||:|
Qy      1 ASYLSTSSSLDY 12


RESULT    3
ID   W27592 standard; peptide; 12 AA.
AC   W27592;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          92.0%;  Score 69;  DB 27;  Length 12;
  Best Local Similarity 83.3%;  Pred. No. 2.00e+00;
  Matches    10;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asflstsssley 12
          ||:|||||||:|
Qy      1 ASYLSTSSSLDY 12


RESULT    4
ID   W27591 standard; peptide; 12 AA.
AC   W27591;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
```

KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
DR N-PSDB; T88404.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 92.0%; Score 69; DB 27; Length 12;
Best Local Similarity 91.7%; Pred. No. 2.00e+00;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 asylstssslhy 12
||||||||||:|
Qy 1 ASYLSTSSSLDY 12

RESULT 5
ID W27569 standard; Protein; 121 AA.
AC W27569;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain variable region.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody;
KW heavy chain; variable region; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
DR N-PSDB; T88404.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 16; Page 76; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain variable region.
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 121 AA;

Query Match 92.0%; Score 69; DB 27; Length 121;
Best Local Similarity 90.9%; Pred. No. 2.00e+00;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 100 sylstasslдy 110
|||||:|||||
Qy 2 SYLSTSSSLDY 12

RESULT 6
ID W27590 standard; peptide; 12 AA.
AC W27590;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).

CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 89.3%; Score 67; DB 27; Length 12;
Best Local Similarity 91.7%; Pred. No. 3.26e+00;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 asylstsfsldy 12
||||||| ||||
Qy 1 ASYLSTSSSLDY 12

RESULT 7
ID W27586 standard; peptide; 12 AA.
AC W27586;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 72; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 84.0%; Score 63; DB 27; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.55e+00;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 asylstsssld 11
|||||||||||
Qy 1 ASYLSTSSSLD 11

RESULT 8
ID W27589 standard; peptide; 12 AA.
AC W27589;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 84.0%; Score 63; DB 27; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.55e+00;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 asylstsssld 11
|||||||||||
Qy 1 ASYLSTSSSLD 11

```
RESULT    9
ID   W27587 standard; peptide; 12 AA.
AC   W27587;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           84.0%;  Score 63;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 8.55e+00;
  Matches   11;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssld 11
          ||||||||||
Qy      1 ASYLSTSSSLD 11


RESULT   10
ID   W27563 standard; peptide; 12 AA.
AC   W27563;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 12
FT                   /label= Tyr, Asn
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 9; Page 65; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           76.0%;  Score 57;  DB 27;  Length 12;
  Best Local Similarity 90.0%;  Pred. No. 3.53e+01;
  Matches    9;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 sylstassld 11
          |||||:||||
Qy      2 SYLSTSSSLD 11


RESULT   11
ID   W27594 standard; peptide; 12 AA.
AC   W27594;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
```

PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Disclosure; Page 75; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 76.0%; Score 57; DB 27; Length 12;
Best Local Similarity 90.0%; Pred. No. 3.53e+01;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 2 sylstassld 11
|||||:||||
Qy 2 SYLSTSSSLD 11

RESULT 12
ID W80222 standard; Protein; 159 AA.
AC W80222;
DT 06-JAN-1999 (first entry)
DE Human succinate-ubiquinone reductase membrane anchor subunit.
KW Human; succinate-ubiquinone reductase membrane anchor subunit; SDHMA;
KW myopathy; progressive external phthalmoplegia; Kearns-Sayre syndrome;
KW myoclonic epilepsy; ecephalopathy; cardiomyopathy; lactic acidosis.
OS Homo sapiens.
PN US5827711-A.
PD 27-OCT-1998.
PF 20-MAR-1997; 828832.
PR 20-MAR-1997; US-828832.
PA (INCY-) INCYTE PHARM INC.
PI Lal P, Shah P;
DR WPI; 98-593998/50.
DR N-PSDB; V66406.
PT DNA encoding succinate-ubiquinone reductase membrane anchor subunit
PT - useful for producing recombinant polypeptide
PS Claim 1; Fig 1A-B; 25pp; English.
CC The present sequence represents a human succinate-ubiquinone reductase
CC membrane anchor subunit (SDHMA). The SDHMA nucleic acid sequence
CC was first identified in Incyte clone 2454416 from the aortic
CC endothelial cell cDNA library. The protein has chemical and structural
CC homology with membrane anchoring subunits from bovine mitochondria,
CC QPs3 and QPs1. The protein also contains potential intramolecular
CC disulphide bridging sites, which are found at Cys11, Cys44, Cys88 and
CC Cys150. SDHMA can be used to treat myopathies, e.g. progressive
CC external phthalmoplegia, Kearns-Sayre syndrome, myoclonic epilepsy,
CC ecephalopathy, cardiomyopathy or lactic acidosis.
SQ Sequence 159 AA;

Query Match 72.0%; Score 54; DB 37; Length 159;
Best Local Similarity 50.0%; Pred. No. 7.08e+01;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 82 aaylnpcsamdy 93
|:||: |::||
Qy 1 ASYLSTSSSLDY 12

RESULT 13
ID W55260 standard; Protein; 159 AA.
AC W55260;
DT 02-JUL-1998 (first entry)
DE H. pylori ORF 06cp30603orf16 protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR N-PSDB; V24669.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claim 14; Page 502; 1145pp; English.
CC This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence 159 AA;

Query Match 72.0%; Score 54; DB 29; Length 159;
Best Local Similarity 60.0%; Pred. No. 7.08e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 60 ylslasaley 69
||| :|:|:|
Qy 3 YLSTSSSLDY 12

RESULT 14
ID W55472 standard; Protein; 177 AA.
AC W55472;
DT 24-JUN-1998 (first entry)
DE H. pylori ORF 06cp30603_23452_c3_80 inner membrane protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN WO9737044-A1.

PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR N-PSDB; V24881.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 14,80; Page 679; 1145pp; English.
CC This sequence is a H. pylori cell envelope inner membrane protein
CC involved in cofactor metabolism.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence 177 AA;

Query Match 72.0%; Score 54; DB 29; Length 177;
Best Local Similarity 60.0%; Pred. No. 7.08e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 60 ylslasaley 69
|||:|:|:|
Qy 3 YLSTSSSLDY 12

RESULT 15
W55688 standard; Protein; 526 AA.
W55688;
DT 07-JUL-1998 (first entry)
DE H. pylori ORF 09cp10713_23452_c3_195 inner membrane protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR N-PSDB; V25097.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 14,80; Pages 947-948; 1145pp; English.
CC This sequence is a H. pylori cell envelope inner membrane protein
CC involved in cofactor metabolism.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence 526 AA;

Query Match 72.0%; Score 54; DB 29; Length 526;
Best Local Similarity 60.0%; Pred. No. 7.08e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 60 ylslasaley 69
|||:|:|:|
Qy 3 YLSTSSSLDY 12

Search completed: Thu Sep 2 12:30:47 1999
Job time : 21 secs.

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:33:13 1999; MasPar time 1.39 Seconds
87.360 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-29
Description: (1-12) from US08599226.pep
Perfect Score: 75
Sequence: 1 ASYLSTSSSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 16.112; Variance 54.206; scale 0.297

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	54	72.0	158	2	US-08-828-	Sequence 3, Applicatio	3.02e+01
2	54	72.0	159	2	US-08-828-	Sequence 1, Applicatio	3.02e+01
3	50	66.7	426	3	PCT-US95-1	Sequence 2, Applicatio	7.56e+01
4	50	66.7	426	1	US-08-336-	Sequence 2, Applicatio	7.56e+01
5	50	66.7	1255	2	US-08-414-	Sequence 68, Applicati	7.56e+01
6	50	66.7	1255	1	US-08-467-	Sequence 68, Applicati	7.56e+01
7	50	66.7	1255	2	US-08-625-	Sequence 2, Applicatio	7.56e+01
8	50	66.7	1255	2	US-08-486-	Sequence 68, Applicati	7.56e+01
9	50	66.7	1255	2	US-08-484-	Sequence 8, Applicatio	7.56e+01
10	50	66.7	1255	2	US-08-468-	Sequence 68, Applicati	7.56e+01
11	50	66.7	1255	2	US-08-356-	Sequence 2, Applicatio	7.56e+01
12	49	65.3	37	3	PCT-US96-0	Sequence 5, Applicatio	9.48e+01
13	49	65.3	117	3	PCT-US96-0	Sequence 11, Applicati	9.48e+01
14	49	65.3	384	1	US-08-457-	Sequence 5, Applicatio	9.48e+01
15	49	65.3	1026	3	PCT-US95-0	Sequence 95, Applicati	9.48e+01
16	49	65.3	1026	1	US-08-453-	Sequence 95, Applicati	9.48e+01
17	49	65.3	1026	1	US-08-453-	Sequence 95, Applicati	9.48e+01
18	49	65.3	1026	1	US-07-998-	Sequence 95, Applicati	9.48e+01
19	49	65.3	1026	2	US-08-268-	Sequence 95, Applicati	9.48e+01
20	49	65.3	1026	3	PCT-US93-1	Sequence 95, Applicati	9.48e+01
21	49	65.3	1203	3	PCT-US95-0	Sequence 103, Applicat	9.48e+01
22	49	65.3	1203	3	PCT-US93-1	Sequence 103, Applicat	9.48e+01
23	49	65.3	1203	2	US-08-268-	Sequence 103, Applicat	9.48e+01
24	49	65.3	1203	1	US-08-453-	Sequence 103, Applicat	9.48e+01
25	49	65.3	1203	1	US-08-453-	Sequence 103, Applicat	9.48e+01
26	49	65.3	1203	1	US-07-998-	Sequence 103, Applicat	9.48e+01
27	48	64.0	459	2	US-08-630-	Sequence 2, Applicatio	1.19e+02
28	48	64.0	459	2	US-08-714-	Sequence 2, Applicati	1.19e+02
29	48	64.0	460	2	US-08-714-	Sequence 7, Applicati	1.19e+02
30	48	64.0	460	2	US-08-630-	Sequence 7, Applicatio	1.19e+02
31	48	64.0	505	2	US-08-936-	Sequence 2, Applicatio	1.19e+02
32	48	64.0	505	2	US-08-630-	Sequence 4, Applicatio	1.19e+02
33	48	64.0	505	2	US-08-936-	Sequence 2, Applicatio	1.19e+02
34	48	64.0	505	2	US-08-714-	Sequence 4, Applicati	1.19e+02
35	48	64.0	505	2	US-08-922-	Sequence 2, Applicati	1.19e+02
36	48	64.0	505	2	US-08-829-	Sequence 2, Applicatio	1.19e+02
37	48	64.0	505	1	US-08-631-	Sequence 2, Applicatio	1.19e+02
38	48	64.0	506	2	US-08-936-	Sequence 8, Applicatio	1.19e+02
39	48	64.0	506	2	US-08-829-	Sequence 8, Applicatio	1.19e+02
40	48	64.0	506	2	US-08-922-	Sequence 8, Applicati	1.19e+02
41	48	64.0	506	2	US-08-936-	Sequence 8, Applicatio	1.19e+02
42	48	64.0	506	1	US-08-631-	Sequence 8, Applicatio	1.19e+02
43	48	64.0	561	2	US-08-714-	Sequence 27, Applicati	1.19e+02
44	47	62.7	290	2	US-08-420-	Sequence 27, Applicati	1.49e+02
45	47	62.7	290	3	PCT-US95-1	Sequence 27, Applicati	1.49e+02

ALIGNMENTS

```
RESULT    1
ID   US-08-828-832-3     STANDARD;     PRT;   158 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08828832
XX
CC    Sequence 3, Application US/08828832
CC    Patent No. 5827711
CC     GENERAL INFORMATION:
CC       APPLICANT:  Lal, Preeti
CC       APPLICANT:  Shah, Purvi
CC       TITLE OF INVENTION:  NOVEL SUCCINATE DEHYDROGENASE SUBUNIT
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/828,832
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0250 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  158 amino acids
CC         TYPE:  amino acid
```

```
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  1575011
SQ   SEQUENCE   158 AA; 17096 MW; 135076 CN;

  Query Match           72.0%;  Score 54;  DB 2;  Length 158;
  Best Local Similarity  50.0%;  Pred. No. 3.02e+01;
  Matches    6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 AAYLNPCSAMDY 92
          |:||:  |::||
Qy      1 ASYLSTSSSLDY 12


RESULT    2
ID   US-08-828-832-1     STANDARD;     PRT;   159 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08828832
XX
CC    Sequence 1, Application US/08828832
CC    Patent No. 5827711
CC     GENERAL INFORMATION:
CC       APPLICANT:  Lal, Preeti
CC       APPLICANT:  Shah, Purvi
CC       TITLE OF INVENTION:  NOVEL SUCCINATE DEHYDROGENASE SUBUNIT
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/828,832
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0250 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  159 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  Consensus
CC         CLONE:  2454416
SQ   SEQUENCE   159 AA; 17043 MW; 129477 CN;

  Query Match           72.0%;  Score 54;  DB 2;  Length 159;
  Best Local Similarity  50.0%;  Pred. No. 3.02e+01;
  Matches    6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     82 AAYLNPCSAMDY 93
          |:||:  |::||
Qy      1 ASYLSTSSSLDY 12


RESULT    3
ID   PCT-US95-13795-2     STANDARD;     PRT;   426 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9513795
XX
CC    Sequence 2, Application PC/TUS9513795
CC     GENERAL INFORMATION:
CC       APPLICANT:  HOLLIS, GREGORY F.
CC       APPLICANT:  PATEL, MAYUR D.
CC       TITLE OF INVENTION:  DNA ENCODING CANINE IMMUNOGLOBULINS
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  126 E. LINCOLN AVENUE; P.O. BOX 2000
CC         CITY:  RAHWAY
CC         STATE:  NEW JERSEY
CC         COUNTRY:  USA
CC         ZIP:  07065-0907
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/13795
CC         FILING DATE:
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  19211Y
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (908) 594-6734
CC         TELEFAX:  (908) 594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  426 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   426 AA; 47234 MW; 1032622 CN;

  Query Match           66.7%;  Score 50;  DB 3;  Length 426;
  Best Local Similarity  63.6%;  Pred. No. 7.56e+01;
  Matches    7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    214 TSYLSPPSPLD 224
          :|||| :|:||
Qy      1 ASYLSTSSSLD 11


RESULT    4
ID   US-08-336-583-2     STANDARD;     PRT;   426 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08336583
XX
CC    Sequence 2, Application US/08336583
CC    Patent No. 5629415
```

```
CC     GENERAL INFORMATION:
CC       APPLICANT:  HOLLIS, GREGORY F.
CC       APPLICANT:  PATEL, MAYUR D.
CC       TITLE OF INVENTION:  DNA ENCODING CANINE IMMUNOGLOBULIN E
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  126 E. LINCOLN AVENUE
CC         CITY:  RAHWAY
CC         STATE:  NEW JERSEY
CC         COUNTRY:  USA
CC         ZIP:  07065-0900
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/336,583
CC         FILING DATE:  09-NOV-1994
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  19211
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (908) 594-6734
CC         TELEFAX:  (908) 594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  426 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   426 AA; 47234 MW; 1032622 CN;

  Query Match           66.7%;  Score 50;  DB 1;  Length 426;
  Best Local Similarity  63.6%;  Pred. No. 7.56e+01;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    214 TSYLSPPSPLD 224
          :|||| :|:||
Qy      1 ASYLSTSSSLD 11


RESULT    5
ID   US-08-414-417B-68     STANDARD;     PRT;   1255 AA.
     xxxxxx
XX
DT
XX
DE   Sequence 68, Application US/08414417B
XX
CC    Sequence 68, Application US/08414417B
CC    Patent No. 5801005
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cheever, Martin A.
CC       APPLICANT:  Disis, Mary L.
CC       TITLE OF INVENTION:  IMMUNE REACTIVITY TO HER-2/neu PROTEIN
CC       TITLE OF INVENTION:  FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
CC       TITLE OF INVENTION:  HER-2/neu ONCOGENE IS ASSOCIATED
CC       NUMBER OF SEQUENCES:  69
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Seed and Berry LLP
CC         STREET:  6300 Columbia Center, 701 Fifth Avenue
CC         CITY:  Seattle
CC         STATE:  Washington
CC         COUNTRY:  US
CC         ZIP:  98104-7092
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/414,417B
CC         FILING DATE:  31-MAR-1995
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Sharkey, Richard G.
CC         REGISTRATION NUMBER:  32,629
CC         REFERENCE/DOCKET NUMBER:  920010.448C2
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (206) 622-4900
CC         TELEFAX:  (206) 682-6031
CC     INFORMATION FOR SEQ ID NO:  68:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1255 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   1255 AA; 137955 MW; 8109851 CN;

  Query Match           66.7%;  Score 50;  DB 2;  Length 1255;
  Best Local Similarity  45.5%;  Pred. No. 7.56e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     63 TYLPTNASLSF 73
          :||:|::|| :
Qy      2 SYLSTSSSLDY 12


RESULT    6
ID   US-08-467-083-68     STANDARD;     PRT;   1255 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 68, Application US/08467083
XX
CC    Sequence 68, Application US/08467083
CC    Patent No. 5726023
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cheever, Martin A.
CC       APPLICANT:  Disis, Mary L.
CC       TITLE OF INVENTION:  IMMUNE REACTIVITY TO HER-2/NEU PROTEIN
CC       TITLE OF INVENTION:  FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH TH
CC       TITLE OF INVENTION:  HER-2/NEU ONCOGENE IS ASSOCIATED
CC       NUMBER OF SEQUENCES:  68
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Seed and Berry
CC         STREET:  6300 Columbia Center, 701 Fifth Avenue
CC         CITY:  Seattle
CC         STATE:  Washington
CC         COUNTRY:  US
CC         ZIP:  98104-7092
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/467,083
CC         FILING DATE:  06-JUN-1995
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/414,417
CC         FILING DATE:  06-JUN-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Sharkey, Richard G.
CC         REGISTRATION NUMBER:  32,629
CC         REFERENCE/DOCKET NUMBER:  920010.448C2
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (206) 622-4900
```

```
CC         TELEFAX:  (206) 682-6031
CC         TELEX:  3723836 SEEDANBERRY
CC     INFORMATION FOR SEQ ID NO:  68:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1255 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   1255 AA; 137955 MW; 8109851 CN;

  Query Match           66.7%;  Score 50;  DB 1;  Length 1255;
  Best Local Similarity  45.5%;  Pred. No. 7.56e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     63 TYLPTNASLSF 73
          :||:|::|| :
Qy      2 SYLSTSSSLDY 12


RESULT    7
ID   US-08-625-101-2     STANDARD;     PRT;   1255 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08625101
XX
CC    Sequence 2, Application US/08625101
CC    Patent No. 5869445
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cheever, Martin A.
CC       APPLICANT:  Disis, Mary L.
CC       TITLE OF INVENTION:  COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE
CC       TITLE OF INVENTION:  REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION
CC       TITLE OF INVENTION:  OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu
CC       TITLE OF INVENTION:  ONCOGENE IS ASSOCIATED
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  SEED and BERRY LLP
CC         STREET:  6300 Columbia Center, 701 Fifth Avenue
CC         CITY:  Seattle
CC         STATE:  Washington
CC         COUNTRY:  USA
CC         ZIP:  98104-7092
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/625,101
CC         FILING DATE:  01-APR-1996
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Sharkey, Richard G.
CC         REGISTRATION NUMBER:  32,629
CC         REFERENCE/DOCKET NUMBER:  920010.448C7
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (206) 622-4900
CC         TELEFAX:  (206) 682-6031
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1255 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1255 AA; 137909 MW; 8111405 CN;

  Query Match           66.7%;  Score 50;  DB 2;  Length 1255;
  Best Local Similarity  45.5%;  Pred. No. 7.56e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     63 TYLPTNASLSF 73
          :||:|::|| :
Qy      2 SYLSTSSSLDY 12


RESULT    8
ID   US-08-486-348A-68     STANDARD;     PRT;   1255 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 68, Application US/08486348A
XX
CC    Sequence 68, Application US/08486348A
CC    Patent No. 5846538
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cheever, Martin A.
CC       APPLICANT:  Disis, Mary L.
CC       TITLE OF INVENTION:  IMMUNE REACTIVITY TO HER-2/neu PROTEIN
CC       TITLE OF INVENTION:  FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH TH
CC       TITLE OF INVENTION:  HER-2/neu ONCOGENE IS ASSOCIATED
CC       NUMBER OF SEQUENCES:  69
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Seed and Berry LLP
CC         STREET:  6300 Columbia Center, 701 Fifth Avenue
CC         CITY:  Seattle
CC         STATE:  Washington
CC         COUNTRY:  US
CC         ZIP:  98104-7092
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/486,348A
CC         FILING DATE:  07-JUN-1995
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Sharkey, Richard G.
CC         REGISTRATION NUMBER:  32,629
CC         REFERENCE/DOCKET NUMBER:  920010.448C6
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (206) 622-4900
CC         TELEFAX:  (206) 682-6031
CC     INFORMATION FOR SEQ ID NO:  68:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1255 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   1255 AA; 137955 MW; 8109851 CN;

  Query Match           66.7%;  Score 50;  DB 2;  Length 1255;
  Best Local Similarity  45.5%;  Pred. No. 7.56e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     63 TYLPTNASLSF 73
          :||:|::|| :
Qy      2 SYLSTSSSLDY 12


RESULT    9
ID   US-08-484-438-8     STANDARD;     PRT;   1255 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/08484438
XX
CC    Sequence 8, Application US/08484438
CC    Patent No. 5811098
CC    Patent No. 5811098 5780031
```

```
CC     GENERAL INFORMATION:
CC       APPLICANT:  Plowman, Gregory D.
CC       APPLICANT:  Culouscou, Jean-Michel
CC       APPLICANT:  Shoyab, Mohammed
CC       APPLICANT:  Siegall, Clay B.
CC       APPLICANT:  Hellstr m, Ingegerd
CC       APPLICANT:  Hellstr m, Karl E.
CC       TITLE OF INVENTION:  HER4 HUMAN RECEPTOR TYROSINE KINASE
CC       NUMBER OF SEQUENCES:  42
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/484,438
CC         FILING DATE:  07-JUN-1995
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/323,442
CC         FILING DATE:  14-OCT-1994
CC         APPLICATION NUMBER:  US 08/150,704
CC         FILING DATE:  10-NOV-1993
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/981,165
CC         FILING DATE:  24-NOV-1992
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Misrock, S. Leslie
CC         REGISTRATION NUMBER:  18,872
CC         REFERENCE/DOCKET NUMBER:  5624-230
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-8864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1255 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1255 AA; 137815 MW; 8105484 CN;

  Query Match           66.7%;  Score 50;  DB 2;  Length 1255;
  Best Local Similarity  45.5%;  Pred. No. 7.56e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db      63 TYLPTNASLSF 73
           :||:|::|| :
Qy       2 SYLSTSSSLDY 12


RESULT   10
ID   US-08-468-545B-68     STANDARD;     PRT;   1255 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 68, Application US/08468545B
XX
CC    Sequence 68, Application US/08468545B
CC    Patent No. 5876712
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cheever, Martin A.
CC       APPLICANT:  Disis, Mary L.
CC       TITLE OF INVENTION:  IMMUNE REACTIVITY TO HER-2/neu PROTEIN
CC       TITLE OF INVENTION:  FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH TH
CC       TITLE OF INVENTION:  HER-2/neu ONCOGENE IS ASSOCIATED
CC       NUMBER OF SEQUENCES:  69
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Seed and Berry LLP
CC         STREET:  6300 Columbia Center, 701 Fifth Avenue
CC         CITY:  Seattle
CC         STATE:  Washington
CC         COUNTRY:  US
CC         ZIP:  98104-7092
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/468,545B
CC         FILING DATE:  06-JUN-1995
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Sharkey, Richard G.
CC         REGISTRATION NUMBER:  32,629
CC         REFERENCE/DOCKET NUMBER:  920010.448C5
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (206) 622-4900
CC         TELEFAX:  (206) 682-6031
CC     INFORMATION FOR SEQ ID NO:  68:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1255 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
SQ   SEQUENCE   1255 AA; 137955 MW; 8109851 CN;

  Query Match           66.7%;  Score 50;  DB 2;  Length 1255;
  Best Local Similarity  45.5%;  Pred. No. 7.56e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db      63 TYLPTNASLSF 73
           :||:|::|| :
Qy       2 SYLSTSSSLDY 12


RESULT   11
ID   US-08-356-786-2     STANDARD;     PRT;   1255 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08356786
XX
CC    Sequence 2, Application US/08356786
CC    Patent No. 5877305
CC     GENERAL INFORMATION:
CC       APPLICANT:  Huston, James S.
CC       APPLICANT:  Oppermann, Hermann
CC       APPLICANT:  Houston, L. L.
CC       APPLICANT:  Ring, David B.
CC       TITLE OF INVENTION:  Biosynthetic Binding Protein for Cancer
CC       TITLE OF INVENTION:  Marker
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Edmund R. Pitcher, Testa, Hurwitz, & Thibeault
CC         STREET:  Exchange Place, 53 State Street
CC         CITY:  Boston
CC         STATE:  Massachusetts
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
```

```
CC         COMPUTER:  IBM PC compatible
CC
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/356,786
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/831,967
CC         FILING DATE:  06-FEB-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Pitcher, Edmund R.
CC         REGISTRATION NUMBER:  27,829
CC         REFERENCE/DOCKET NUMBER:  CRP-053
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617) 248-7000
CC         TELEFAX:  (617) 248-7100
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1255 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1255 AA; 137909 MW; 8111405 CN;

  Query Match           66.7%;  Score 50;  DB 2;  Length 1255;
  Best Local Similarity  45.5%;  Pred. No. 7.56e+01;
  Matches     5;  Conservative     5;  Mismatches   1;  Indels   0;  Gaps   0;

Db      63 TYLPTNASLSF 73
           :||:|::|| :
Qy       2 SYLSTSSSLDY 12


RESULT   12
ID   PCT-US96-08730-5     STANDARD;     PRT;   37 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application PC/TUS9608730
XX
CC    Sequence 5, Application PC/TUS9608730
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cassels, Frederick
CC       APPLICANT:  Anderson, Jeffrey
CC       APPLICANT:  Carter, John Mark
CC       TITLE OF INVENTION:  Methods of Raising Antibodies Against E.
CC       TITLE OF INVENTION:  Coli of the Family CSF-CFA./1
CC       NUMBER OF SEQUENCES:  15
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Glenna Hendricks
CC         STREET:  P.O. Box 2509
CC         CITY:  Fairfax
CC         STATE:  VA
CC         COUNTRY:  USA
CC         ZIP:  22031
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US96/08730
CC         FILING DATE:  03-JUN-1996
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Hendricks, Glenna
CC         REGISTRATION NUMBER:  32,535
CC         REFERENCE/DOCKET NUMBER:  PCT/US96/08730
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703) 591-4470
CC         TELEFAX:  (703) 591-4428
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  37 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  internal
SQ   SEQUENCE   37 AA; 3864 MW; 7776 CN;

  Query Match           65.3%;  Score 49;  DB 3;  Length 37;
  Best Local Similarity  50.0%;  Pred. No. 9.48e+01;
  Matches     6;  Conservative     3;  Mismatches   3;  Indels   0;  Gaps   0;

Db      22 GSYLPTAVELTY 33
           :|||:|:  | |
Qy       1 ASYLSTSSSLDY 12


RESULT   13
ID   PCT-US96-08730-11     STANDARD;     PRT;   117 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 11, Application PC/TUS9608730
XX
CC    Sequence 11, Application PC/TUS9608730
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cassels, Frederick
CC       APPLICANT:  Anderson, Jeffrey
CC       APPLICANT:  Carter, John Mark
CC       TITLE OF INVENTION:  Methods of Raising Antibodies Against E.
CC       TITLE OF INVENTION:  Coli of the Family CSF-CFA./1
CC       NUMBER OF SEQUENCES:  15
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Glenna Hendricks
CC         STREET:  P.O. Box 2509
CC         CITY:  Fairfax
CC         STATE:  VA
CC         COUNTRY:  USA
CC         ZIP:  22031
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US96/08730
CC         FILING DATE:  03-JUN-1996
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Hendricks, Glenna
CC         REGISTRATION NUMBER:  32,535
CC         REFERENCE/DOCKET NUMBER:  PCT/US96/08730
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703) 591-4470
CC         TELEFAX:  (703) 591-4428
CC     INFORMATION FOR SEQ ID NO:  11:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  117 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  internal
SQ   SEQUENCE   117 AA; 12389 MW; 76297 CN;
```

```
  Query Match          65.3%;  Score 49;  DB 3;  Length 117;
  Best Local Similarity  50.0%;  Pred. No. 9.48e+01;
  Matches     6;  Conservative    3;  Mismatches   3;  Indels   0;  Gaps   0;

Db      22 GSYLPTAVELTY 33
           :|||:|:  | |
Qy       1 ASYLSTSSSLDY 12


RESULT   14
ID   US-08-457-245-5     STANDARD;     PRT;   384 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/08457245
XX
     Sequence 5, Application US/08457245
     Patent No. 5573915
      GENERAL INFORMATION:
CC      APPLICANT:  BARRY III, Clifton E.
CC      APPLICANT:  YUAN, Ying
CC      TITLE OF INVENTION:  CLONING AND EXPRESSION OF DNA INVOLVED
CC      TITLE OF INVENTION:  IN THE BIOSYNTHESIS OF CYCLOPROPANATED MYCOLIC ACIDS IN
CC      TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS
CC      NUMBER OF SEQUENCES:  21
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/457,245
CC        FILING DATE:
CC        CLASSIFICATION:  435
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Chambers, Guy W.
CC        REGISTRATION NUMBER:  30,617
CC        REFERENCE/DOCKET NUMBER:  15280-216000
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC    INFORMATION FOR SEQ ID NO:  5:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  384 amino acids
CC        TYPE:  amino acid
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  ORF2 protein
SQ   SEQUENCE   384 AA; 41963 MW; 701271 CN;

  Query Match          65.3%;  Score 49;  DB 1;  Length 384;
  Best Local Similarity  54.5%;  Pred. No. 9.48e+01;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db     235 SSYLPTKAALD 245
           :|||:| ::||
Qy       1 ASYLSTSSSLD 11


RESULT   15
ID   PCT-US95-08071-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application PC/TUS9508071
XX
CC    Sequence 95, Application PC/TUS9508071
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  6300 Sears Tower, 233 S. Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/08071
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/12588
CC         FILING DATE:  23 DEC 1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/998,003
CC         FILING DATE:  29 DEC 1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Noland, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32149
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match          65.3%;  Score 49;  DB 3;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 9.48e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db     423 FLQTTTPLDY 432
           :| |:::|||
Qy       3 YLSTSSSLDY 12


Search completed: Thu Sep  2 12:33:21 1999
Job time : 8 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

n on: Thu Sep 2 12:31:04 1999; MasPar time 3.15 Seconds
152.820 Million cell updates/sec

bular output not generated.

```
Title:           >US-08-599-226-29
Description:     (1-12) from US08599226.pep
Perfect Score:   75
Sequence:        1 ASYLSTSSSLDY 12

Scoring table:   PAM 150
                 Gap 15

Searched:        122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:        pir60
                   1:pir1 2:pir2 3:pir3 4:pir4

Statistics:      Mean 24.557;  Variance 30.670;  scale 0.801
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	54	72.0	503	2	F64713	protein-export membra	2.65e+00
2	54	72.0	526	2	D71805	protein-export membra	2.65e+00
3	53	70.7	265	2	S38380	Hroxl protein - Calif	4.22e+00
4	52	69.3	917	1	ACGAE	glutamate receptor pr	6.67e+00
5	51	68.0	471	2	C71439	hypothetical protein	1.05e+01
6	51	68.0	1069	2	T00040	BH-protocadherin PCDH	1.05e+01
7	51	68.0	1069	2	T00043	BH-protocadherin-a -	1.05e+01
8	51	68.0	1072	2	T00041	BH-protocadherin PCDH	1.05e+01
9	51	68.0	1200	2	T00042	BH-protocadherin PCDH	1.05e+01
10	50	66.7	293	2	D70108	conserved hypothetica	1.64e+01
11	50	66.7	469	2	D70048	ABC transporter (amin	1.64e+01
12	50	66.7	626	1	IKECB	colicin Ib - Escheric	1.64e+01
13	50	66.7	989	2	I56333	apolipoprotein B - ra	1.64e+01
14	50	66.7	1186	2	T03180	tyrosine protein kina	1.64e+01
15	50	66.7	1255	1	A24571	protein-tyrosine kina	1.64e+01
16	49	65.3	105	2	S69755	hypothetical protein	2.54e+01
17	49	65.3	112	4	S59333	hypothetical protein	2.54e+01
18	49	65.3	377	2	F71520	hypothetical protein	2.54e+01
19	49	65.3	392	2	B70242	conserved hypothetica	2.54e+01
20	49	65.3	409	2	G64677	NADH dehydrogenase (u	2.54e+01
21	49	65.3	409	2	E71838	nadh oxidoreductase I	2.54e+01
22	49	65.3	564	2	A38271	serotonin receptor 7	2.54e+01
23	49	65.3	650	2	F70974	probable acrA1 protei	2.54e+01
24	49	65.3	1469	2	A55095	chromosome condensati	2.54e+01
25	48	64.0	272	2	D64155	hypothetical protein	3.92e+01
26	48	64.0	297	2	G70708	probable purC protein	3.92e+01
27	48	64.0	325	2	S75747	hypothetical protein	3.92e+01
28	48	64.0	428	2	A55044	beta-4C-adrenergic re	3.92e+01
29	48	64.0	441	2	G70822	probable secY protein	3.92e+01
30	48	64.0	459	2	S68519	tub protein, testis -	3.92e+01
31	48	64.0	475	2	S55093	hypothetical protein	3.92e+01
32	48	64.0	477	2	F71918	hypothetical protein	3.92e+01
33	48	64.0	505	2	S68518	tub protein, brain -	3.92e+01
34	48	64.0	814	2	I39627	nicotine dehydrogenas	3.92e+01
35	48	64.0	954	2	G71496	hypothetical protein	3.92e+01
36	48	64.0	1211	2	S54500	alpha,alpha-trehalase	3.92e+01
37	47	62.7	179	2	S65534	light-harvesting chlo	5.99e+01
38	47	62.7	386	2	F70231	conserved hypothetica	5.99e+01
39	47	62.7	593	2	S49525	glycoprotein G - simi	5.99e+01
40	47	62.7	626	2	C25035	colicin Ia - Escheric	5.99e+01
41	47	62.7	684	2	G70744	hypothetical protein	5.99e+01
42	47	62.7	826	2	B36203	iron-responsive eleme	5.99e+01
43	47	62.7	949	3	T03030	hypothetical protein	5.99e+01
44	47	62.7	963	2	A57238	iron-responsive eleme	5.99e+01
45	47	62.7	976	2	PC4208	valine--tRNA ligase (	5.99e+01

ALIGNMENTS

```
RESULT    1
ENTRY            F64713     #type complete
TITLE            protein-export membrane protein - Helicobacter pylori (strain
                   26695)
ORGANISM         #formal_name Helicobacter pylori
DATE             09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
                   12-Feb-1999
ACCESSIONS       F64713
REFERENCE        A64520
   #authors      Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
                   Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
                   H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
                   J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
                   Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
                   McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
                   Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
                   Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
                   Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
                   W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
                   C.M.; Venter, J.C.
   #journal      Nature (1997) 388:539-547
   #title        The complete genome sequence of the gastric pathogen
                   Helicobacter pylori.
   #cross-references MUID:97394467
   #accession    F64713
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-503 ##label TOM
      ##cross-references GB:AE000652; GB:AE000511; NID:g2314720; PID:g2314730;
                        TIGR:HP1550
GENETICS
   #start_codon  GTG
CLASSIFICATION   #superfamily protein export membrane protein secD
SUMMARY          #length 503  #molecular-weight 54247  #checksum 3320

  Query Match             72.0%;  Score 54;  DB 2;  Length 503;
  Best Local Similarity  60.0%;  Pred. No. 2.65e+00;
  Matches     6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Db       38 YLSLASALEY 47
            |||  :|:|:|
Qy        3 YLSTSSSLDY 12


RESULT    2
ENTRY            D71805     #type complete
```

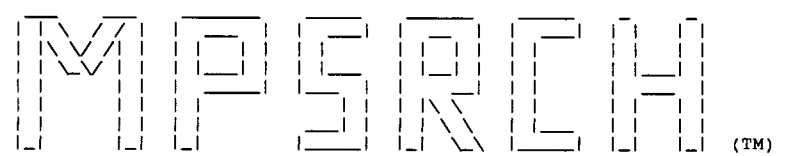
```
TITLE           protein-export membrane protein - Helicobacter pylori (strain
                  J99)
ORGANISM        #formal_name Helicobacter pylori
   #variety     strain J99
DATE            12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                  05-Mar-1999
ACCESSIONS      D71805
REFERENCE       A71800
   #authors     Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                  Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                  B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                  Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                  Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                  G.F.; Trust, T.J.
   #journal     Nature (1999) 397:176-180
   #title       Genomic sequence comparison of two unrelated isolates of the
                  human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession   D71805
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-526 ##label ARN
      ##cross-references GB:AE001567; GB:AE001439; NID:g4156065; PID:g4156069
      ##experimental_source strain J99
GENETICS
   #gene        secD
CLASSIFICATION  #superfamily protein export membrane protein secD
SUMMARY         #length 526  #molecular-weight 56796  #checksum 5813

  Query Match           72.0%;  Score 54;  DB 2;  Length 526;
  Best Local Similarity  60.0%;  Pred. No. 2.65e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     60 YLSLASALEY 69
          ||| :|:|:|
Qy      3 YLSTSSSLDY 12


RESULT    3
ENTRY           S38380     #type complete
TITLE           Hrox1 protein - California red abalone
ORGANISM        #formal_name Haliotis rufescens #common_name California red
                  abalone
DATE            13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
                  17-Oct-1997
ACCESSIONS      S38380
REFERENCE       S38380
   #authors     Degnan, B.M.; Degnan, S.M.; Morse, D.E.
   #submission  submitted to the EMBL Data Library, September 1993
   #description Expression of Hrox, a gastropod mollusc Hox homeobox gene, is
                  progressively restricted during morphogenesis from
                  trochophore to veliger larval forms.
   #accession   S38380
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-265 ##label DEG
      ##cross-references EMBL:X75217; NID:g407414; PID:g407415
CLASSIFICATION  #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS        DNA binding; homeobox; nucleus; transcription regulation
FEATURE
   164-220          #domain homeobox homology #label HOX
SUMMARY         #length 265  #molecular-weight 29579  #checksum 8996

  Query Match           70.7%;  Score 53;  DB 2;  Length 265;
  Best Local Similarity  75.0%;  Pred. No. 4.22e+00;
  Matches     9;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 GSYLSMSSSKDY 92
          :|||| ||| ||
Qy      1 ASYLSTSSSLDY 12


RESULT    4
ENTRY           ACGAE      #type complete
TITLE           glutamate receptor precursor - great pond snail
ORGANISM        #formal_name Lymnaea stagnalis #common_name great pond snail
DATE            31-Mar-1992 #sequence_revision 28-Oct-1994 #text_change
                  05-Sep-1997
ACCESSIONS      S18443; S15681
REFERENCE       S18443
   #authors     Hutton, M.L.; Harvey, R.J.; Barnard, E.A.; Darlison, M.G.
   #journal     FEBS Lett. (1991) 292:111-114
   #title       Cloning of a cDNA that encodes an invertebrate glutamate
                  receptor subunit.
   #cross-references MUID:92070466
   #accession   S18443
      ##status        nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-917 ##label HUT
      ##cross-references EMBL:X60086
REFERENCE       S15681
   #authors     Hutton, M.L.; Bhandal, N.S.; Harvey, R.J.; Usherwood, P.N.R.;
                  Darlison, M.G.
   #submission  submitted to the EMBL Data Library, June 1991
   #description PCR-mediated cloning of a cDNA that encodes a functional
                  molluscan glutamate receptor subunit.
   #accession   S15681
      ##molecule_type mRNA
      ##residues      1-362,'K',364-776,'S',778-845,'R',847-886,'S',888-917
                       ##label HU2
      ##cross-references EMBL:X60086; NID:g9628; PID:g9629
CLASSIFICATION  #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS        glycoprotein; ion channel; neurotransmitter receptor;
                  transmembrane protein
FEATURE
   1-19             #domain signal sequence #status predicted #label SIG\
   20-917           #product glutamate receptor #status predicted #label
                      MAT\
   429-853          #domain glutamate receptor homology #label GRH\
   559-578          #domain transmembrane #status predicted #label TM1\
   599-617          #domain transmembrane #status predicted #label TM2\
   628-646          #domain transmembrane #status predicted #label TM3\
   819-839          #domain transmembrane #status predicted #label TM4\
   62,95,121,125,229,
   251,261,272,418,
   419,424,491,881  #binding_site carbohydrate (Asn) (covalent) #status
                      predicted
SUMMARY         #length 917  #molecular-weight 103106  #checksum 8952

  Query Match           69.3%;  Score 52;  DB 1;  Length 917;
  Best Local Similarity  54.5%;  Pred. No. 6.67e+00;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    729 AYLTESSTIDY 739
          :||: ||::||
Qy      2 SYLSTSSSLDY 12


RESULT    5
ENTRY           C71439     #type complete
TITLE           hypothetical protein - Arabidopsis thaliana
ORGANISM        #formal_name Arabidopsis thaliana #common_name mouse-ear
                  cress
   #variety     columbia
DATE            03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
                  05-Dec-1998
ACCESSIONS      C71439
REFERENCE       A71400
   #authors     Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.;
                  Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.;
                  Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel,
                  K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.;
                  Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.;
                  Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.;
                  Lecharny, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.;
                  Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger,
```

```
                  M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
                  James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka,
                  A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.;
                  Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.;
                  Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.;
                  Rechman, S.; Ansorge, W.; Cooke, R.; Berger, C.; Delseny,
                  M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
                  Schueller, C.; Chalwatzis, N.
   #journal       Nature (1998) 391:485-488
   #title         Analysis of 1.9 Mb of contiguous sequence from chromosome 4
                    of Arabidopsis thaliana.
   #cross-references MUID:98121113
   #accession     C71439
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-471 ##label BEV
      ##cross-references GB:Z97342; NID:g2245031; PID:e327038; PID:g2245065
GENETICS
   #map_position 4COP9-4G3845
SUMMARY           #length 471  #molecular-weight 52785  #checksum 7455

  Query Match            68.0%;  Score 51;  DB 2;  Length 471;
  Best Local Similarity  63.6%;  Pred. No. 1.05e+01;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db      19 AGYLTSSSDLD 29
           |:||::|| ||
Qy       1 ASYLSTSSSLD 11


RESULT    6
ENTRY             T00040     #type complete
TITLE             BH-protocadherin PCDH7 - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                    22-Jan-1999
ACCESSIONS        T00040
REFERENCE         Z14074
   #authors       Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
                    Sugano, S.
   #journal       Genomics (1998) 49:458-461
   #title         Cloning, expression analysis, and chromosomal localization of
                    BH-protocadherin (PCDH7), a novel member of the cadherin
                    superfamily.
   #accession     T00040
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-1069 ##label YOS
      ##cross-references EMBL:AB006755; NID:d1184677; PID:d1026122
      ##experimental_source clone BH-Pcdh-a
GENETICS
   #map_position 4p15
SUMMARY           #length 1069  #molecular-weight 116104  #checksum 9974

  Query Match            68.0%;  Score 51;  DB 2;  Length 1069;
  Best Local Similarity  60.0%;  Pred. No. 1.05e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     486 FLHTSTPLDY 495
           :| ||::|||
Qy       3 YLSTSSSLDY 12


RESULT    7
ENTRY             T00043     #type complete
TITLE             BH-protocadherin-a - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                    22-Jan-1999
ACCESSIONS        T00043
REFERENCE         Z14075
   #authors       Yoshida, K.
   #submission    submitted to the EMBL Data Library, August 1997
   #accession     T00043
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-1069 ##label YOS
      ##cross-references EMBL:AB006758; NID:d1227200; PID:d1033562
GENETICS
   #gene          pcdh7
   #map_position 5C3-D
SUMMARY           #length 1069  #molecular-weight 116313  #checksum 4821

  Query Match            68.0%;  Score 51;  DB 2;  Length 1069;
  Best Local Similarity  60.0%;  Pred. No. 1.05e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     486 FLHTSAPLDY 495
           :| ||::|||
Qy       3 YLSTSSSLDY 12


RESULT    8
ENTRY             T00041     #type complete
TITLE             BH-protocadherin PCDH7 (clone BH-Pcdh-b) - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                    22-Jan-1999
ACCESSIONS        T00041
REFERENCE         Z14074
   #authors       Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
                    Sugano, S.
   #journal       Genomics (1998) 49:458-461
   #title         Cloning, expression analysis, and chromosomal localization of
                    BH-protocadherin (PCDH7), a novel member of the cadherin
                    superfamily.
   #accession     T00041
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-1072 ##label YOS
      ##cross-references EMBL:AB006756; NID:d1184678; PID:d1026123
      ##experimental_source clone BH-Pcdh-b
GENETICS
   #map_position 4p15
SUMMARY           #length 1072  #molecular-weight 116462  #checksum 9727

  Query Match            68.0%;  Score 51;  DB 2;  Length 1072;
  Best Local Similarity  60.0%;  Pred. No. 1.05e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     486 FLHTSTPLDY 495
           :| ||::|||
Qy       3 YLSTSSSLDY 12


RESULT    9
ENTRY             T00042     #type complete
TITLE             BH-protocadherin PCDH7 (clone BH-Pcdh-c) - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                    22-Jan-1999
ACCESSIONS        T00042
REFERENCE         Z14074
   #authors       Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
                    Sugano, S.
   #journal       Genomics (1998) 49:458-461
   #title         Cloning, expression analysis, and chromosomal localization of
                    BH-protocadherin (PCDH7), a novel member of the cadherin
                    superfamily.
   #accession     T00042
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-1200 ##label YOS
      ##cross-references EMBL:AB006757; NID:d1184679; PID:d1026124
      ##experimental_source clone BH-Pcdh-c
```

```
GENETICS
   #map_position 4p15
SUMMARY          #length 1200  #molecular-weight 130337  #checksum 7152

  Query Match            68.0%;  Score 51;  DB 2;  Length 1200;
  Best Local Similarity  60.0%;  Pred. No. 1.05e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    439 FLHTSTPLDY 448
          :| ||::|||
Qy      3 YLSTSSSLDY 12


RESULT   10
ENTRY            D70108     #type complete
TITLE            conserved hypothetical protein BB0068 - Lyme disease
                   spirochete
ORGANISM         #formal_name Borrelia burgdorferi #common_name Lyme disease
                   spirochete
DATE             13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                   05-Jun-1998
ACCESSIONS       D70108
REFERENCE        A70100
   #authors      Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
                   Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
                   Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
                   J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
                   Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
                   Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
                   J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
                   Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
                   K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
   #journal      Nature (1997) 390:580-586
   #title        Genomic sequence of a Lyme disease spirochaete, Borrelia
                   burgdorferi.
   #cross-references MUID:98065943
   #accession    D70108
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-293 ##label KLE
      ##cross-references GB:AE001120; GB:AE000783; NID:g2687951; PID:g2687956;
                          TIGR:BB0068
      ##experimental_source strain B31
SUMMARY          #length 293  #molecular-weight 33278  #checksum 5223

  Query Match            66.7%;  Score 50;  DB 2;  Length 293;
  Best Local Similarity  60.0%;  Pred. No. 1.64e+01;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    198 AYLSTPNSLE 207
          :|||||::||:
Qy      2 SYLSTSSSLD 11


RESULT   11
ENTRY            D70048     #type complete
TITLE            ABC transporter (amino acid permease) homolog yvsH - Bacillus
                   subtilis
ORGANISM         #formal_name Bacillus subtilis
DATE             05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                   24-Sep-1998
ACCESSIONS       D70048
REFERENCE        A69580
   #authors      Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                   Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                   Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                   A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                   Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                   Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                   Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                   Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                   Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                   M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                   S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                   Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                   C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                   Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                   Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                   Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                   Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                   Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                   Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                   M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                   M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                   V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                   A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                   Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                   Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                   Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                   Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                   B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                   Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                   Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                   Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                   Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                   Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                   K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                   Yoshikawa, H.; Danchin, A.
   #journal      Nature (1997) 390:249-256
   #title        The complete genome sequence of the Gram-positive bacterium
                   Bacillus subtilis.
   #cross-references MUID:98044033
   #accession    D70048
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-469 ##label KUN
      ##cross-references GB:Z99121; GB:AL009126; NID:g2635827; PID:e1186022;
                          PID:g2635847
      ##experimental_source strain 168
GENETICS
   #gene         yvsH
CLASSIFICATION   #superfamily L-lysine transport protein
SUMMARY          #length 469  #molecular-weight 50258  #checksum 4200

  Query Match            66.7%;  Score 50;  DB 2;  Length 469;
  Best Local Similarity  36.4%;  Pred. No. 1.64e+01;
  Matches     4;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db    376 TFLTTAATLAY 386
          ::|:|:::| |
Qy      2 SYLSTSSSLDY 12


RESULT   12
ENTRY            IKECB      #type complete
TITLE            colicin Ib - Escherichia coli plasmid ColIb
ORGANISM         #formal_name Escherichia coli
DATE             17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change
                   20-Mar-1998
ACCESSIONS       A93533; D25035; A22503; B22503; A03503
REFERENCE        A93533
   #authors      Varley, J.M.; Boulnois, G.J.
   #journal      Nucleic Acids Res. (1984) 12:6727-6739
   #title        Analysis of a cloned colicin IB gene: complete nucleotide
                   sequence and implications for regulation of expression.
   #cross-references MUID:85014128
   #accession    A93533
      ##molecule_type DNA
      ##residues      1-626 ##label VAR
      ##cross-references GB:X01009; NID:g41141; PID:g41142
      ##note          the authors translated the codon GAA for residue 294 as
                       Gln
REFERENCE        A91822
   #authors      Mankovich, J.A.; Hsu, C.H.; Konisky, J.
```

```
   #journal      J. Bacteriol. (1986) 168:228-236
   #title        DNA and amino acid sequence analysis of structural and
                   immunity genes of colicins Ia and Ib.
   #cross-references MUID:87008385
   #accession    D25035
      ##molecule_type DNA
      ##residues      1-626 ##label MAN
REFERENCE        A22503
   #authors      Mankovich, J.A.; Lai, P.H.; Gokul, N.; Konisky, J.
   #journal      J. Biol. Chem. (1984) 259:8764-8768
   #title        Organization of the colicin Ib gene.
   #cross-references MUID:84264487
   #accession    A22503
      ##molecule_type DNA
      ##residues      1-40 ##label MA2
   #accession    B22503
      ##molecule_type protein
      ##residues      2-21 ##label MA3
REFERENCE        A93546
   #authors      Varley, J.M.; Boulnois, G.J.
   #journal      Nucleic Acids Res. (1984) 12:8748
   #contents     annotation; corrigendum
COMMENT     This bactericidal protein functions by depolarizing the cytoplasmic
              membrane of sensitive cells.
GENETICS
   #gene         cib
   #genome       plasmid
CLASSIFICATION   #superfamily colicin IB
KEYWORDS         antibiotic; bacteriocin; toxin; transmembrane protein
FEATURE
   2-626              #product colicin Ib #status experimental #label MAT
SUMMARY          #length 626  #molecular-weight 69923  #checksum 8113

  Query Match           66.7%;  Score 50;  DB 1;  Length 626;
  Best Local Similarity  63.6%;  Pred. No. 1.64e+01;
  Matches     7;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    333 NYLLTHSGLDY 343
          :|| | |:|||
Qy      2 SYLSTSSSLDY 12


RESULT   13
ENTRY            I56333     #type fragment
TITLE            apolipoprotein B - rat (fragment)
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
                   23-Feb-1997
ACCESSIONS       I56333
REFERENCE        I56333
   #authors      Reuben, M.A.; Svenson, K.L.; Doolittle, M.H.; Johnson, D.F.;
                   Lusis, A.J.; Elovson, J.
   #journal      J. Lipid Res. (1988) 29:1337-1347
   #title        Biosynthetic relationships between three rat apolipoprotein B
                   peptides.
   #cross-references MUID:89176719
   #accession    I56333
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-989 ##label RES
      ##cross-references GB:M27440; NID:g623548; PID:g623549
GENETICS
   #gene         apoB
CLASSIFICATION   #superfamily apolipoprotein B
SUMMARY          #length 989  #checksum 1918

  Query Match           66.7%;  Score 50;  DB 2;  Length 989;
  Best Local Similarity  60.0%;  Pred. No. 1.64e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    175 YLQASTSLHY 184
          || :|:|| |
Qy      3 YLSTSSSLDY 12


RESULT   14
ENTRY            T03180     #type complete
TITLE            tyrosine protein kinase homolog - Chilo iridescent virus
ORGANISM         #formal_name Chilo iridescent virus
DATE             24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
                   24-Mar-1999
ACCESSIONS       T03180
REFERENCE        Z14834
   #authors      Bahr, U.; Tidona, C.A.; Darai, G.
   #journal      Virus Genes (1997) 15:235-245
   #title        The DNA sequence of Chilo iridescent virus between the genome
                   coordinates 0.101 and 0.391; similarities in coding
                   strategy between insect and vertebrate iridoviruses.
   #accession    T03180
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-1186 ##label BAH
      ##cross-references EMBL:AF003534; NID:g2738385; PID:g2738451
SUMMARY          #length 1186  #molecular-weight 138020  #checksum 6252

  Query Match           66.7%;  Score 50;  DB 2;  Length 1186;
  Best Local Similarity  63.6%;  Pred. No. 1.64e+01;
  Matches     7;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Db    503 TYLSRESILDY 513
          :|||  | |||
Qy      2 SYLSTSSSLDY 12


RESULT   15
ENTRY            A24571     #type complete
TITLE            protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor -
                   human
ALTERNATE_NAMES  c-erb-B-2 protein precursor; kinase-related transforming
                   protein erbB2; v-erbB-related protein HER-2/neu
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change
                   22-May-1998
ACCESSIONS       A24571; A25491; A44188; B44188; I59509; I57622
REFERENCE        A24571
   #authors      Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.;
                   Miyajima, N.; Saito, T.; Toyoshima, K.
   #journal      Nature (1986) 319:230-234
   #title        Similarity of protein encoded by the human c-erb-B-2 gene to
                   epidermal growth factor receptor.
   #cross-references MUID:86118663
   #accession    A24571
      ##molecule_type mRNA
      ##residues      1-1255 ##label YAM
      ##cross-references GB:X03363; NID:g31197; PID:g31198
REFERENCE        A25491
   #authors      Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1985) 82:6497-6501
   #title        A v-erbB-related protooncogene, c-erbB-2, is distinct from
                   the c-erbB-1/epidermal growth factor-receptor gene and is
                   amplified in a human salivary adenocarcinoma.
   #cross-references MUID:86016729
   #accession    A25491
      ##molecule_type DNA
      ##residues      737-1031 ##label SEM
      ##cross-references GB:M11767; NID:g182163; PID:g553282
REFERENCE        A44188
   #authors      Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray,
                   A.; McGrath, J.; Seeburg, P.H.; Libermann, T.A.;
                   Schlessinger, J.; Francke, U.; Levinson, A.; Ullrich, A.
   #journal      Science (1985) 230:1132-1139
   #title        Tyrosine kinase receptor with extensive homology to EGF
                   receptor shares chromosomal location with neu oncogene.
   #cross-references MUID:86070181
   #accession    A44188
      ##molecule_type DNA
```

```
      ##residues      740-910 ##label COU1
      ##cross-references GB:M12036; NID:g183988; PID:g183989
   #accession    B44188
      ##molecule_type mRNA
      ##residues      1-517,'RALL',522,'S',524-654,'V',656-1169,'A',1171-1255
                       ##label COU2
      ##cross-references GB:M11730; NID:g183986
REFERENCE        I59509
   #authors      King, C.R.; Kraus, M.H.; Aaronson, S.A.
   #journal      Science (1985) 229:974-976
   #title        Amplification of a novel v-erbB-related gene in a human
                  mammary carcinoma.
   #cross-references MUID:85272597
   #accession    I59509
      ##status        translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      832-909 ##label REX
      ##cross-references GB:L29395; NID:g459807; PID:g459808
REFERENCE        I57622
   #authors      Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger,
                  J.; Givol, D.
   #journal      Mol. Cell. Biol. (1987) 7:2597-2601
   #title        Human HER2 (neu) promoter: evidence for multiple mechanisms
                  for transcriptional initiation.
   #cross-references MUID:87286898
   #accession    I57622
      ##status        translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-191 ##label TAL
      ##cross-references GB:M16792; NID:g183983; PID:g553332
COMMENT    Amplification and overexpression of this erbB-related gene occurs
             in about 30% of human breast and ovarian cancers.
GENETICS
   #gene         GDB:ERBB2; NGL; NEU; HER-2
      ##cross-references GDB:120613; OMIM:164870
   #map_position 17q21.1-17q21.1
   #introns      25/1; 75/3; 147/1; 883/3
   #note         the list of introns is incomplete
FUNCTION
   #description  catalyzes the phosphorylation of a peptidyl tyrosine residue
                   by ATP
CLASSIFICATION   #superfamily epidermal growth factor receptor; protein kinase
                   homology
KEYWORDS         ATP; autophosphorylation; duplication; glycoprotein;
                   phosphoprotein; phosphotransferase; proto-oncogene;
                   receptor; transforming protein; transmembrane protein;
                   tyrosine-specific protein kinase

FEATURE
   1-21             #domain signal sequence #status predicted #label SIG\
   22-1255          #product protein-tyrosine kinase erbB2 #status predicted
                      #label MAT\
   22-653           #domain extracellular #status predicted #label EXT\
   70-304           #domain EGF receptor extracellular domain repeat #label
                      EE1\
   395-605          #domain EGF receptor extracellular domain repeat #label
                      EE2\
   654-675          #domain transmembrane #status predicted #label TMM\
   676-1255         #domain intracellular #status predicted #label INT\
   718-983          #domain protein kinase homology #label KIN\
   726-734          #region protein kinase ATP-binding motif\
   68,124,187,259,530,
   571,629          #binding_site carbohydrate (Asn) (covalent) #status
                      predicted\
   686              #binding_site phosphate (Thr) (covalent) (by protein
                      kinase C) #status predicted\
   753              #active_site Lys #status predicted\
   1139,1221,1222,
   1248             #binding_site phosphate (Tyr) (covalent) (by
                      autophosphorylation) #status predicted
SUMMARY          #length 1255  #molecular-weight 137909  #checksum 9382

  Query Match           66.7%;  Score 50;  DB 1;  Length 1255;
  Best Local Similarity 45.5%;  Pred. No. 1.64e+01;

  Matches    5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db      63 TYLPTNASLSF 73
           :||:|::|| :
Qy       2 SYLSTSSSLDY 12


Search completed: Thu Sep  2 12:31:23 1999
Job time : 19 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:31:39 1999; MasPar time 2.20 Seconds
154.507 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-29
Description: (1-12) from US08599226.pep
Perfect Score: 75
Sequence: 1 ASYLSTSSSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 25.063; Variance 28.009; scale 0.895

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	54	72.0	158	1	DHSD_BOVIN	SUCCINATE DEHYDROGENAS	9.05e-01
2	54	72.0	159	1	DHSD_HUMAN	SUCCINATE DEHYDROGENAS	9.05e-01
3	52	69.3	917	1	GLRK_LYMST	GLUTAMATE RECEPTOR PRE	2.48e+00
4	51	68.0	1038	1	SOG_DROME	DORSAL-VENTRAL PATTERN	4.06e+00
5	50	66.7	626	1	CEIB_ECOLI	COLICIN IB PROTEIN.	6.61e+00
6	50	66.7	1255	1	ERB2_HUMAN	ERBB-2 RECEPTOR PROTEI	6.61e+00
7	49	65.3	564	1	5HT1_DROME	5-HYDROXYTRYPTAMINE RE	1.07e+01
8	49	65.3	1469	1	DP27_CAEEL	CHROMOSOME CONDENSATIO	1.07e+01
9	48	64.0	37	1	YRYL_CAEEL	HYPOTHETICAL 4.1 KD PR	1.71e+01
10	48	64.0	272	1	YIGL_HAEIN	HYPOTHETICAL PROTEIN H	1.71e+01
11	48	64.0	297	1	PUR7_MYCLE	PHOSPHORIBOSYLAMINOIMI	1.71e+01
12	48	64.0	297	1	PUR7_MYCTU	PHOSPHORIBOSYLAMINOIMI	1.71e+01
13	48	64.0	428	1	B4AR_MELGA	BETA-4C ADRENERGIC REC	1.71e+01
14	48	64.0	441	1	SECY_MYCTU	PREPROTEIN TRANSLOCASE	1.71e+01
15	48	64.0	475	1	YM61_YEAST	HYPOTHETICAL 55.3 KD P	1.71e+01
16	48	64.0	505	1	TUB_MOUSE	TUBBY PROTEIN.	1.71e+01
17	48	64.0	506	1	TUB_HUMAN	TUBBY PROTEIN HOMOLOG.	1.71e+01
18	48	64.0	697	1	TRP_SCHPO	TRYPTOPHAN SYNTHASE (E	1.71e+01
19	48	64.0	1211	1	ATH1_YEAST	VACUOLAR ACID TREHALAS	1.71e+01
20	47	62.7	330	1	ODBA_BACSU	2-OXOISOVALERATE DEHYD	2.71e+01
21	47	62.7	517	1	DMPN_PSESP	PHENOL HYDROXYLASE P3	2.71e+01
22	47	62.7	579	1	YR47_CAEEL	HYPOTHETICAL 66.0 KD P	2.71e+01
23	47	62.7	626	1	CEIA_ECOLI	COLICIN IA PROTEIN.	2.71e+01
24	47	62.7	684	1	YV18_MYCTU	HYPOTHETICAL 57.3 KD P	2.71e+01
25	47	62.7	687	1	VIUA_VIBCH	VIBRIOBACTIN RECEPTOR	2.71e+01
26	47	62.7	780	1	YAD3_SCHPO	HYPOTHETICAL 88.2 KD P	2.71e+01
27	47	62.7	848	1	AMPN_LACLA	AMINOPEPTIDASE N (EC 3	2.71e+01
28	47	62.7	963	1	IRE2_RAT	IRON-RESPONSIVE ELEMEN	2.71e+01
29	47	62.7	963	1	IRE2_HUMAN	IRON-RESPONSIVE ELEMEN	2.71e+01
30	47	62.7	1294	1	YA3B_SCHPO	HYPOTHETICAL 149.2 KD	2.71e+01
31	47	62.7	1522	1	YB6A_SCHPO	HYPOTHETICAL 171.5 KD	2.71e+01
32	46	61.3	127	1	MBP_RAT	MYELIN BASIC PROTEIN S	4.26e+01
33	46	61.3	132	1	YW07_MYCTU	HYPOTHETICAL 14.7 KD P	4.26e+01
34	46	61.3	167	1	MBP_CAVPO	MYELIN BASIC PROTEIN (	4.26e+01
35	46	61.3	171	1	MBP_PANTR	MYELIN BASIC PROTEIN (	4.26e+01
36	46	61.3	196	1	MBP_HUMAN	MYELIN BASIC PROTEIN (	4.26e+01
37	46	61.3	352	1	YDH3_SCHPO	HYPOTHETICAL 39.7 KD P	4.26e+01
38	46	61.3	442	1	HPPD_DAUCA	4-HYDROXYPHENYLPYRUVAT	4.26e+01
39	46	61.3	464	1	WCAM_ECOLI	COLANIC ACID BIOSYNTHE	4.26e+01
40	46	61.3	471	1	UFO1_MAIZE	FLAVONOL 3-O-GLUCOSYLT	4.26e+01
41	46	61.3	471	1	UFO3_MAIZE	FLAVONOL 3-O-GLUCOSYLT	4.26e+01
42	46	61.3	471	1	UFO2_MAIZE	FLAVONOL 3-O-GLUCOSYLT	4.26e+01
43	46	61.3	509	1	STK_HYDAT	TYROSINE-PROTEIN KINAS	4.26e+01
44	46	61.3	577	1	PYRH_YEAST	CTP SYNTHASE 2 (EC 6.3	4.26e+01
45	46	61.3	653	1	YJU4_YEAST	HYPOTHETICAL 75.9 KD P	4.26e+01

ALIGNMENTS

```
RESULT    1
ID   DHSD_BOVIN     STANDARD;      PRT;   158 AA.
AC   Q95123;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT
DE   PRECURSOR (CYBS) (SUCCINATE-UBIQUINONE REDUCTASE MEMBRANE ANCHOR
DE   SUBUNIT) (QPS3).
GN   SDHD OR SDH4.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RA   SHENOY S.K., YU L., YU C.A.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBUNIT: COMPOSED OF A 27 KD IRON PROTEIN (IP), A 70 KD
CC       FLAVOPROTEIN (FP) AND A CYTOCHROME B COMPOSED OF TWO INTEGRAL
CC       MEMBRANE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U50987; G1575011; -.
KW   TRICARBOXYLIC ACID CYCLE; ELECTRON TRANSPORT; HEME; TRANSMEMBRANE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   TRANSIT       1     55       MITOCHONDRION (POTENTIAL).
FT   CHAIN        56    158       SUCCINATE DEHYDROGENASE [UBIQUINONE]
FT                                 CYTOCHROME B SMALL SUBUNIT.
FT   TRANSMEM     70     90       POTENTIAL.
FT   TRANSMEM    125    141       POTENTIAL.
SQ   SEQUENCE   158 AA;  17096 MW;  703D5238 CRC32;

  Query Match           72.0%;  Score 54;  DB 1;  Length 158;
  Best Local Similarity 50.0%;  Pred. No. 9.05e-01;
  Matches    6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 AAYLNPCSAMDY 92
```

```
          |:||:  |::||
Qy      1 ASYLSTSSSLDY 12


RESULT    2
ID   DHSD_HUMAN     STANDARD;      PRT;   159 AA.
AC   O14521;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT
DE   PRECURSOR (CYBS) (SUCCINATE-UBIQUINONE REDUCTASE MEMBRANE ANCHOR
DE   SUBUNIT).
GN   SDHD OR SDH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 98194224.
RA   HIRAWAKE H., TANIWAKI M., KIJIMA S., KITA K.;
RT   "Cytochrome b in human complex II (succinate-ubiquinone
RT   oxidoreductase): cDNA cloning of the components in liver mitochondria
RT   and chromosome assignment of the genes for the large (SDHC) and small
RT   (SDHD) subunits to 1q21 and 11q23.";
RL   CYTOGENET. CELL GENET. 79:132-138(1997).
CC   -!- SUBUNIT: COMPOSED OF A 27 KD IRON PROTEIN (IP), A 70 KD
CC       FLAVOPROTEIN (FP) AND A CYTOCHROME B COMPOSED OF TWO INTEGRAL
CC       MEMBRANE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB006202; D1022913; -.
DR   MIM; 602690; -.
KW   TRICARBOXYLIC ACID CYCLE; ELECTRON TRANSPORT; HEME; TRANSMEMBRANE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   TRANSIT       1     56       MITOCHONDRION (POTENTIAL).
FT   CHAIN        57    159       SUCCINATE DEHYDROGENASE [UBIQUINONE]
FT                                 CYTOCHROME B SMALL SUBUNIT.
FT   TRANSMEM     71     91       POTENTIAL.
FT   TRANSMEM    126    142       POTENTIAL.
SQ   SEQUENCE   159 AA;  17043 MW;  F4221825 CRC32;

  Query Match           72.0%;  Score 54;  DB 1;  Length 159;
  Best Local Similarity  50.0%;  Pred. No. 9.05e-01;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     82 AAYLNPCSAMDY 93
          |:||:  |::||
Qy      1 ASYLSTSSSLDY 12


RESULT    3
ID   GLRK_LYMST     STANDARD;      PRT;   917 AA.
AC   P26591;
DT   01-AUG-1992 (REL. 23, CREATED)
DT   01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT   01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE   GLUTAMATE RECEPTOR PRECURSOR.
OS   LYMNAEA STAGNALIS (GREAT POND SNAIL).
OC   EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PULMONATA; BASOMMATOPHORA;
OC   LYMNAEIDAE; LYMNAEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92070466.
RA   HUTTON M.L., HARVEY R.J., BARNARD E.A., DARLISON M.G.;
RT   "Cloning of a cDNA that encodes an invertebrate glutamate receptor
RT   subunit.";
RL   FEBS LETT. 292:111-114(1991).
CC   -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT
CC       MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC
CC       ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60086; G9629; -.
DR   PIR; S15681; ACGAE.
DR   PIR; S18443; S18443.
DR   PFAM; PF00060; lig_chan; 1.
KW   RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
KW   TRANSMEMBRANE.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    917       GLUTAMATE RECEPTOR.
FT   DOMAIN       21    558       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    559    578       POTENTIAL.
FT   TRANSMEM    599    617       POTENTIAL.
FT   TRANSMEM    628    646       POTENTIAL.
FT   TRANSMEM    819    839       POTENTIAL.
FT   CARBOHYD     62     62       POTENTIAL.
FT   CARBOHYD     95     95       POTENTIAL.
FT   CARBOHYD    121    121       POTENTIAL.
FT   CARBOHYD    125    125       POTENTIAL.
FT   CARBOHYD    229    229       POTENTIAL.
FT   CARBOHYD    251    251       POTENTIAL.
FT   CARBOHYD    261    261       POTENTIAL.
FT   CARBOHYD    272    272       POTENTIAL.
FT   CARBOHYD    418    418       POTENTIAL.
FT   CARBOHYD    419    419       POTENTIAL.
FT   CARBOHYD    424    424       POTENTIAL.
FT   CARBOHYD    491    491       POTENTIAL.
FT   CARBOHYD    775    775       POTENTIAL.
FT   CARBOHYD    881    881       POTENTIAL.
SQ   SEQUENCE   917 AA;  103139 MW;  879CBEDC CRC32;

  Query Match           69.3%;  Score 52;  DB 1;  Length 917;
  Best Local Similarity  54.5%;  Pred. No. 2.48e+00;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    729 AYLTESSTIDY 739
          :||: ||::||
Qy      2 SYLSTSSSLDY 12


RESULT    4
ID   SOG_DROME      STANDARD;      PRT;  1038 AA.
AC   Q24025;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   DORSAL-VENTRAL PATTERNING PROTEIN SOG (SHORT GASTRULATION PROTEIN).
GN   SOG.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95047398.
RA   FRANCOIS V., SOLLOWAY M., O'NEILL J.W., EMERY J., BIER E.;
```

```
RT   "Dorsal-ventral patterning of the Drosophila embryo depends on a
RT   putative negative growth factor encoded by the short gastrulation
RT   gene.";
RL   GENES DEV. 8:2602-2616(1994).
CC   -!- FUNCTION: PUTATIVE NEGATIVE GROWTH FACTOR; ANTAGONIST OF DPP, A
CC       PROTEIN INVOLVED IN PATTERNING THE DORSAL REGION AND IN THE
CC       DEVELOPMENT OF THE NEUROECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: ABUTS THE DORSAL DPP-EXPRESSING CELLS IN A
CC       LATERAL STRIPE 14-16 CELLS WIDE. LATER IN EMBRYOGENESIS IT IS
CC       EXPRESSED IN NEUROECTODERM AND IN THE ENDODERM SPACED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS OF THE DEVELOPING GUT.
CC   -!- DEVELOPMENTAL STAGE: EMBRYOGENESIS.
CC   -!- SIMILARITY: TO XENOPUS DORSALIZING FACTOR CHORDIN.
CC   -!- SIMILARITY: CONTAINS 4 VWFC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18774; G1203794; -.
DR   FLYBASE; FBgn0003463; sog.
DR   PROSITE; PS01208; VWFC; 2.
DR   PFAM; PF00093; vwc; 4.
KW   TRANSMEMBRANE; DEVELOPMENTAL PROTEIN; REPEAT; GROWTH FACTOR;
KW   GROWTH REGULATION; SIGNAL-ANCHOR.
FT   DOMAIN        1     53       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     54     74       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       75   1038       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      100    175       VWFC 1.
FT   REPEAT      421    522       SR1.
FT   REPEAT      592    668       SR2.
FT   REPEAT      677    754       SR3.
FT   DOMAIN      742    804       VWFC 2.
FT   DOMAIN      830    899       VWFC 3.
FT   DOMAIN      939   1020       VWFC 4.
FT   CARBOHYD    179    179       POTENTIAL.
FT   CARBOHYD    287    287       POTENTIAL.
FT   CARBOHYD    520    520       POTENTIAL.
FT   CARBOHYD    666    666       POTENTIAL.
FT   CARBOHYD    752    752       POTENTIAL.
FT   CARBOHYD    821    821       POTENTIAL.
SQ   SEQUENCE   1038 AA;  115514 MW;  DC4DAFF5 CRC32;

  Query Match           68.0%;  Score 51;  DB 1;  Length 1038;
  Best Local Similarity  60.0%;  Pred. No. 4.06e+00;
  Matches    6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    497 YLNTDGSLAY 506
          ||:| :|| |
Qy      3 YLSTSSSLDY 12


RESULT   5
ID   CEIB_ECOLI     STANDARD;      PRT;   626 AA.
AC   P04479;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   COLICIN IB PROTEIN.
GN   CIB.
OS   ESCHERICHIA COLI.
OG   PLASMID INCI1 COLIB-P9.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ESCHERICHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87008385.
RA   MANKOVICH J.A., HSU C.-H., KONISKY J.;
RT   "DNA and amino acid sequence analysis of structural and immunity
RT   genes of colicins Ia and Ib.";
RL   J. BACTERIOL. 168:228-236(1986).
RN   [2]
RP   SEQUENCE OF 1-40 FROM N.A.
RX   MEDLINE; 84264487.
RA   MANKOVICH J.A., LAI P.H., GOKUL N., KONISKY J.;
RT   "Organization of the colicin Ib gene. Promoter structure and immunity
RT   domain.";
RL   J. BIOL. CHEM. 259:8764-8768(1984).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85014128.
RA   VARLEY J.M., BOULNOIS G.J.;
RT   "Analysis of a cloned colicin Ib gene: complete nucleotide sequence
RT   and implications for regulation of expression.";
RL   NUCLEIC ACIDS RES. 12:6727-6739(1984).
RN   [4]
RP   ERRATUM.
RA   VARLEY J.M., BOULNOIS G.J.;
RL   NUCLEIC ACIDS RES. 12:8748-8748(1984).
CC   -!- FUNCTION: THIS COLICIN IS A CHANNEL-FORMING COLICIN. THIS CLASS OF
CC       TRANSMEMBRANE TOXINS DEPOLARIZE THE CYTOPLASMIC MEMBRANE, LEADING
CC       TO DISSIPATION OF CELLULAR ENERGY.
CC   -!- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
CC       AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
CC   -!- SIMILARITY: BELONGS TO THE CHANNEL FORMING COLICIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02071; G144640; -.
DR   EMBL; X01009; G41142; -.
DR   EMBL; M13820; G144648; -.
DR   PIR; A03503; IKECB.
DR   PIR; A22503; A22503.
DR   PIR; D25035; D25035.
DR   PROSITE; PS00276; CHANNEL_COLICIN; 1.
DR   PFAM; PF01024; Colicin; 1.
DR   HSSP; P06716; 1CII.
KW   PLASMID; BACTERIOCIN; COLICIN; TOXIN; TRANSMEMBRANE.
FT   TRANSMEM    588    612       POTENTIAL.
SQ   SEQUENCE   626 AA;  69923 MW;  983D11B0 CRC32;

  Query Match           66.7%;  Score 50;  DB 1;  Length 626;
  Best Local Similarity  63.6%;  Pred. No. 6.61e+00;
  Matches    7;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    333 NYLLTHSGLDY 343
          :|| | |:|||
Qy      2 SYLSTSSSLDY 12


RESULT   6
ID   ERB2_HUMAN     STANDARD;      PRT;  1255 AA.
AC   P04626;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   ERBB-2 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
DE   (P185ERBB2) (NEU PROTO-ONCOGENE) (C-ERBB-2).
GN   ERBB2 OR HER2 OR NGL OR NEU.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
```

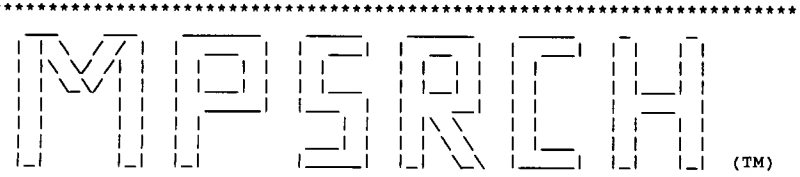
```
RX   MEDLINE; 86118663.
RA   YAMAMOTO T., IKAWA S., AKIYAMA T., SEMBA K., NOMURA N., MIYAJIMA N.,
RA   SAITO T., TOYOSHIMA K.;
RT   "Similarity of protein encoded by the human c-erb-B-2 gene to
RT   epidermal growth factor receptor.";
RL   NATURE 319:230-234(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86070181.
RA   COUSSENS L., YANG-FENG T.L., LIAO Y.C., CHEN E., GRAY A.,
RA   MCGRATH J., SEEBURG P.H., LIBERMANN T.A., SCHLESSINGER J.,
RA   FRANCKE U., LEVINSON A., ULLRICH A.;
RT   "Tyrosine kinase receptor with extensive homology to EGF receptor
RT   shares chromosomal location with neu oncogene.";
RL   SCIENCE 230:1132-1139(1985).
RN   [3]
RP   SEQUENCE OF 737-1031 FROM N.A.
RX   MEDLINE; 86016729.
RA   SEMBA K., KAMATA N., TOYOSHIMA K., YAMAMOTO T.;
RT   "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT   c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT   human salivary gland adenocarcinoma.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 82:6497-6501(1985).
CC   -!- FUNCTION: NEUREGULINS AND GP30 ARE POTENTIAL LIGANDS FOR THIS
CC       RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC       (POTENTIAL).
CC   -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC       RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11767; G553282; -.
DR   EMBL; M11761; G553282; JOINED.
DR   EMBL; M11762; G553282; JOINED.
DR   EMBL; M11763; G553282; JOINED.
DR   EMBL; M11764; G553282; JOINED.
DR   EMBL; M11765; G553282; JOINED.
DR   EMBL; M11766; G553282; JOINED.
DR   EMBL; M11730; G306840; -.
DR   EMBL; M12036; G183989; -.
DR   EMBL; X03363; G31198; -.
DR   PIR; A25491; A25491.
DR   PIR; A24571; A24571.
DR   MIM; 164870; -.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00757; Furin-like; 1.
DR   PFAM; PF01030; Recep_L_domain; 2.
DR   HSSP; P11362; 1FGI.
KW   TRANSMEMBRANE; GLYCOPROTEIN; MULTIGENE FAMILY; RECEPTOR; SIGNAL;
KW   TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; PHOSPHORYLATION.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   1255       ERBB2 PROTEIN.
FT   DOMAIN      22    652       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   653    675       POTENTIAL.
FT   DOMAIN     676   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     720    987       PROTEIN KINASE.
FT   NP_BIND    726    734       ATP (BY SIMILARITY).
FT   BINDING    753    753       ATP (BY SIMILARITY).
FT   ACT_SITE   845    845       BY SIMILARITY.
FT   MOD_RES   1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES   1248   1248       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD    68     68       POTENTIAL.
FT   CARBOHYD   124    124       POTENTIAL.
FT   CARBOHYD   187    187       POTENTIAL.
FT   CARBOHYD   259    259       POTENTIAL.
FT   CARBOHYD   530    530       POTENTIAL.
FT   CARBOHYD   571    571       POTENTIAL.
FT   CARBOHYD   629    629       POTENTIAL.
FT   CONFLICT   655    655       I -> V (IN REF. 2).
FT   CONFLICT  1170   1170       P -> A (IN REF. 2).
SQ   SEQUENCE  1255 AA;  137909 MW;  715C377C CRC32;

  Query Match           66.7%;  Score 50;  DB 1;  Length 1255;
  Best Local Similarity 45.5%;  Pred. No. 6.61e+00;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     63 TYLPTNASLSF 73
          :||:|::|| :
Qy      2 SYLSTSSSLDY 12


RESULT    7
ID   5HT1_DROME     STANDARD;      PRT;   564 AA.
AC   P20905;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   5-HYDROXYTRYPTAMINE RECEPTOR 1 (5-HT RECEPTOR) (SEROTONIN RECEPTOR).
GN   5HT-R1 OR 5-HT7.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=HEAD;
RX   MEDLINE; 91062395.
RA   WITZ P., AMLAIKY N., PLASSAT J.-L., MAROTEAUX L., BORRELLI E., HEN R.;
RT   "Cloning and characterization of a Drosophila serotonin receptor that
RT   activates adenylate cyclase.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 87:8940-8944(1990).
CC   -!- FUNCTION: THIS IS ONE OF THE SEVERAL DIFFERENT RECEPTORS FOR
CC       5-HYDROXYTRYPTAMINE (SEROTONIN), A BIOGENIC HORMONE THAT FUNCTION
CC       AS A NEUROTRANSMITTER, A HORMONE, AND A MITOGEN. THE ACTIVITY OF
CC       THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLATE
CC       CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: HEAD.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       BUT WITH ONE EXTRA POTENTIAL TRANSMEMBRANE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M55533; G156725; -.
DR   PIR; A38271; A38271.
DR   GCRDB; GCR_0023; -.
DR   FLYBASE; FBgn0004573; 5-HT7.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT.
FT   TRANSMEM     29     51       0 (POTENTIAL).
FT   TRANSMEM    165    188       1 (POTENTIAL).
FT   DOMAIN      189    198       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    199    222       2 (POTENTIAL).
FT   DOMAIN      223    236       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    237    258       3 (POTENTIAL).
```

```
FT   DOMAIN      259    278       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    279    302       4 (POTENTIAL).
FT   DOMAIN      303    330       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    331    353       5 (POTENTIAL).
FT   DOMAIN      354    454       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    455    476       6 (POTENTIAL).
FT   DOMAIN      477    487       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    488    510       7 (POTENTIAL).
FT   DOMAIN      511    564       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       89    106       9 X 2 AA TANDEM REPEATS OF G-S.
FT   DISULFID    235    314       BY SIMILARITY.
SQ   SEQUENCE   564 AA;  60861 MW;  312369E8 CRC32;

  Query Match           65.3%;  Score 49;  DB 1;  Length 564;
  Best Local Similarity  40.0%;  Pred. No. 1.07e+01;
  Matches     4;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    261 YLAITKPLEY 270
          ||: : :|:|
Qy      3 YLSTSSSLDY 12


RESULT    8
ID   DP27_CAEEL     STANDARD;      PRT;  1469 AA.
AC   P48996;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   CHROMOSOME CONDENSATION PROTEIN DPY-27.
GN   DPY-27 OR R13G10.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 95042743.
RA   CHUANG P.-T., ALBERTSON D.G., MEYER B.J.;
RT   "DPY-27:a chromosome condensation protein homolog that regulates C.
RT   elegans dosage compensation through association with the X
RT   chromosome.";
RL   CELL 79:459-474(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   GARDINER A.;
RL   SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: ACTS TO REDUCE EXPRESSION OF BOTH HERMAPHRODITE X
CC       CHROMOSOMES. DPY-27 BECOMES SPECIFICALLY LOCALIZED TO THE X
CC       CHROMOSOMES OF WILD-TYPE XX EMBRYOS, BUT REMAINS DIFFUSELY
CC       DISTRIBUTED THROUGHOUT THE NUCLEI OF MALE (XO) EMBRYOS. COULD
CC       IMPLEMENTS DOSAGE COMPENSATION BY CONDENSING THE CHROMATIN
CC       STRUCTURE OF X IN A MANNER THAT CAUSES GENERAL REDUCTION OF X
CC       CHROMOSOME EXPRESSION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN EMBRYOS AND EARLY-STAGED LARVAE.
CC   -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC       FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L35274; G529385; -.
DR   EMBL; Z35602; E1348662; -.
DR   WORMPEP; R13G10.1; CE01052.
KW   ATP-BINDING; COILED COIL; NUCLEAR PROTEIN.
FT   NP_BIND     122    129       ATP (POTENTIAL).
FT   DOMAIN      875    916       COILED COIL (POTENTIAL).
FT   DOMAIN     1014   1171       COILED COIL (POTENTIAL).
FT   DOMAIN     1253   1280       ALA/ASP-RICH (DA-BOX).
FT   MUTAGEN     128    128       K->E: LOSS OF FUNCTION.
FT   MUTAGEN     128    128       K->I: LOSS OF FUNCTION.
SQ   SEQUENCE   1469 AA;  169618 MW;  DD8782B5 CRC32;

  Query Match           65.3%;  Score 49;  DB 1;  Length 1469;
  Best Local Similarity  45.5%;  Pred. No. 1.07e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    316 AYLNTFNNLNY 326
          :||:| ::|:|
Qy      2 SYLSTSSSLDY 12


RESULT    9
ID   YRYL_CAEEL     STANDARD;      PRT;    37 AA.
AC   Q19177;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 4.1 KD PROTEIN F07H5.4 IN CHROMOSOME II.
GN   F07H5.4.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   STEWARD C.;
RL   SUBMITTED (DEC-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   REVISIONS.
RC   STRAIN=BRISTOL N2;
RA   JONES S.J.M.;
RL   SUBMITTED (MAR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z68314; E1345182; -.
DR   WORMPEP; F07H5.4; CE09231.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   37 AA;  4082 MW;  C16B5292 CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 37;
  Best Local Similarity  60.0%;  Pred. No. 1.71e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db      2 YLDDPSSIDY 11
            ||  :||:||
Qy      3 YLSTSSSLDY 12


RESULT   10
ID   YIGL_HAEIN     STANDARD;      PRT;   272 AA.
AC   P44771;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL PROTEIN HI0597.
GN   HI0597.
OS   HAEMOPHILUS INFLUENZAE.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC   HAEMOPHILUS.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   SCIENCE 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO THE COF/YBHA/YIDA/YIGL (E.COLI) / YCSE/YXEH
CC       (B.SUBTILIS) FAMILY. STRONG, TO E.COLI YIGL.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32741; G1573586; -.
DR   TIGR; HI0597; -.
DR   PROSITE; PS01228; COF_1; 1.
DR   PROSITE; PS01229; COF_2; 1.
DR   PFAM; PF00592; DUF3; 1.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   272 AA;  30523 MW;  7F53B65C CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 272;
  Best Local Similarity  58.3%;  Pred. No. 1.71e+01;
  Matches     7;  Conservative    1;  Mismatches   4;  Indels   0;  Gaps   0;

Db    261 ARYLRTQFGLDY 272
          | || |  :|||
Qy      1 ASYLSTSSSLDY 12


RESULT   11
ID   PUR7_MYCLE     STANDARD;      PRT;   297 AA.
AC   O08361;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PHOSPHORIBOSYLAMINOIMIDAZOLE-SUCCINOCARBOXAMIDE SYNTHASE (EC 6.3.2.6)
DE   (SAICAR SYNTHETASE).
GN   PURC OR MLCB5.16.
OS   MYCOBACTERIUM LEPRAE.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BADCOCK K., CHURCHER C.M., PARKHILL J., BARRELL B.G.,
RA   RAJANDREAM M.A.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: ATP + 1-(5-PHOSPHORIBOSYL)-4-CARBOXY-5-
CC       AMINOIMIDAZOLE + L-ASPARTATE = ADP + ORTHOPHOSPHATE + 1-(5-
CC       PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)-5-AMINOIMIDAZOLE.
CC   -!- PATHWAY: SEVENTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER SAICAR SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z95151; E315054; -.
DR   PROSITE; PS01057; SAICAR_SYNTHETASE_1; 1.
DR   PROSITE; PS01058; SAICAR_SYNTHETASE_2; 1.
DR   PFAM; PF01259; SAICAR_synt; 1.
KW   PURINE BIOSYNTHESIS; LIGASE.
SQ   SEQUENCE   297 AA;  33163 MW;  56F60A12 CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 297;
  Best Local Similarity  54.5%;  Pred. No. 1.71e+01;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    104 GYLTGSGLLDY 114
          :||: |: |||
Qy      2 SYLSTSSSLDY 12


RESULT   12
ID   PUR7_MYCTU     STANDARD;      PRT;   297 AA.
AC   Q59566; P77904;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PHOSPHORIBOSYLAMINOIMIDAZOLE-SUCCINOCARBOXAMIDE SYNTHASE (EC 6.3.2.6)
DE   (SAICAR SYNTHETASE).
GN   PURC OR MTCY369.24.
OS   MYCOBACTERIUM TUBERCULOSIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96425868.
RA   JACKSON M., BERTHET F.-X., OTAL I., RAUZIER J., MARTIN C.,
RA   GICQUEL B., GUILHOT C.;
RT   "The Mycobacterium tuberculosis purine biosynthetic pathway:
RT   isolation and characterization of the purC and purL genes.";
RL   MICROBIOLOGY 142:2439-2447(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   MCLEAN J., HARRIS D., BARRELL B.G., RAJANDREAM M.A.;
RL   SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: ATP + 1-(5-PHOSPHORIBOSYL)-4-CARBOXY-5-
CC       AMINOIMIDAZOLE + L-ASPARTATE = ADP + ORTHOPHOSPHATE + 1-(5-
CC       PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)-5-AMINOIMIDAZOLE.
CC   -!- PATHWAY: SEVENTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER SAICAR SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z80226; E266570; -.
DR   EMBL; U34957; G1144522; -.
DR   PROSITE; PS01057; SAICAR_SYNTHETASE_1; 1.
DR   PROSITE; PS01058; SAICAR_SYNTHETASE_2; 1.
DR   PFAM; PF01259; SAICAR_synt; 1.
KW   PURINE BIOSYNTHESIS; LIGASE.
FT   CONFLICT    124    124       A -> R (IN REF. 1).
FT   CONFLICT    166    166       A -> P (IN REF. 1).
SQ   SEQUENCE   297 AA;  32930 MW;  23A860DC CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 297;
  Best Local Similarity  54.5%;  Pred. No. 1.71e+01;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    104 GYLTGSGLLDY 114
```

```
          :||: |: |||
Qy      2 SYLSTSSSLDY 12


RESULT   13
ID   B4AR_MELGA     STANDARD;      PRT;   428 AA.
AC   P43141;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   BETA-4C ADRENERGIC RECEPTOR.
GN   ADRB4C.
OS   MELEAGRIS GALLOPAVO (COMMON TURKEY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; MELEAGRIDIDAE; MELEAGRIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95014249.
RA   CHEN X.-H., HARDEN T.K., NICHOLAS R.A.;
"Molecular cloning and characterization of a novel beta-adrenergic
receptor.";
RL   J. BIOL. CHEM. 269:24810-24819(1994).
CC   -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC       INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC       PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: BROAD TISSUE DISTRIBUTION.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U13977; G556604; -.
DR   EMBL; U13978; G555882; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   HSSP; P07700; 1DEP.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT   DOMAIN       1     25       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     26    49       1 (POTENTIAL).
FT   DOMAIN       50    58       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     59    77       2 (POTENTIAL).
FT   DOMAIN       78    97       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     98    119      3 (POTENTIAL).
FT   DOMAIN       120   141      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     142   164      4 (POTENTIAL).
FT   DOMAIN       165   189      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     190   211      5 (POTENTIAL).
FT   DOMAIN       212   261      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     262   283      6 (POTENTIAL).
FT   DOMAIN       284   294      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     295   315      7 (POTENTIAL).
FT   DOMAIN       316   428      CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     8     8        POTENTIAL.
FT   CARBOHYD     13    13       POTENTIAL.
FT   DISULFID     96    175      BY SIMILARITY.
FT   LIPID        329   329      PALMITATE (BY SIMILARITY).
SQ   SEQUENCE   428 AA;  47398 MW;  8B794F0C CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 428;
  Best Local Similarity  40.0%;  Pred. No. 1.71e+01;
  Matches     4;  Conservative     5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    122 YLAITAPLQY 131
          ||: :::|:|
Qy      3 YLSTSSSLDY 12


RESULT   14
ID   SECY_MYCTU     STANDARD;      PRT;   441 AA.
AC   P94926;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PREPROTEIN TRANSLOCASE SECY SUBUNIT.
GN   SECY OR RV0732 OR MTV041.06.
OS   MYCOBACTERIUM TUBERCULOSIS, AND MYCOBACTERIUM BOVIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.TUBERCULOSIS; STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   NATURE 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.BOVIS; STRAIN=BCG;
RA   KIM J.K., CHOE Y.K.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECA AND SECE
CC       TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,
CC       BY FORMING PART OF A CHANNEL (BY SIMILARITY).
CC   -!- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
CC       COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE SECY/SEC61-ALPHA FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL021958; E1253270; -.
DR   EMBL; U77912; G1813627; -.
DR   PROSITE; PS00755; SECY_1; 1.
DR   PROSITE; PS00756; SECY_2; 1.
DR   PFAM; PF00344; secY; 1.
KW   PROTEIN TRANSPORT; TRANSMEMBRANE; TRANSLOCATION.
FT   TRANSMEM     18    38       POTENTIAL.
FT   TRANSMEM     57    77       POTENTIAL.
FT   TRANSMEM     78    98       POTENTIAL.
FT   TRANSMEM     124   144      POTENTIAL.
FT   TRANSMEM     157   177      POTENTIAL.
FT   TRANSMEM     180   200      POTENTIAL.
FT   TRANSMEM     215   235      POTENTIAL.
FT   TRANSMEM     272   292      POTENTIAL.
FT   TRANSMEM     318   338      POTENTIAL.
FT   TRANSMEM     382   402      POTENTIAL.
FT   TRANSMEM     404   424      POTENTIAL.
SQ   SEQUENCE   441 AA;  47611 MW;  26E1FAC6 CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 441;
  Best Local Similarity  50.0%;  Pred. No. 1.71e+01;
  Matches     6;  Conservative     4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    309 GTYLSDPSNLVY 320
```

```
          ::||| :|:| |
Qy      1 ASYLSTSSSLDY 12


RESULT   15
ID   YM61_YEAST     STANDARD;      PRT;   475 AA.
AC   Q03652;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 55.3 KD PROTEIN IN RAR1-SCJ1 INTERGENIC REGION.
GN   YMR211W OR YM8261.05.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   DEDMAN K., BROWN D., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.,
RA   WALSH S.V.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49809; G854463; -.
DR   PFAM; PF00091; tubulin; 1.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   475 AA;  55312 MW;  18FA4F03 CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 475;
  Best Local Similarity 50.0%;  Pred. No. 1.71e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    321 YLTTAITLGY 330
          ||:|: :| |
Qy      3 YLSTSSSLDY 12


Search completed: Thu Sep  2 12:31:47 1999
Job time : 8 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:32:06 1999; MasPar time 4.45 Seconds
147.067 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-29
Description: (1-12) from US08599226.pep
Perfect Score: 75
Sequence: 1 ASYLSTSSSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.003; Variance 31.514; scale 0.762

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	55	73.3	649	3	O60167	PROTEIN COMPLEX ASSEMB	2.58e+00
2	54	72.0	349	5	O17959	M01B2.5 PROTEIN.	4.07e+00
3	54	72.0	503	2	O26074	PROTEIN-EXPORT MEMBRAN	4.07e+00
4	53	70.7	127	14	O11696	NUCLEOPROTEIN (FRAGMEN	6.38e+00
5	53	70.7	265	5	Q25144	HROX1.	6.38e+00
6	53	70.7	897	5	Q17336	LET 858.	6.38e+00
7	53	70.7	1238	5	O61198	F15E6.6 PROTEIN.	6.38e+00
8	52	69.3	372	5	P91143	SIMILAR TO ACETYLTRANS	9.96e+00
9	52	69.3	475	13	O93514	AXIAL PROTOCADHERIN (F	9.96e+00
10	52	69.3	932	5	O01623	SIMILAR TO LIGAND-GATE	9.96e+00
11	51	68.0	83	9	O48385	ORF83.	1.55e+01
12	51	68.0	471	10	O23552	HYPOTHETICAL 52.8 KD P	1.55e+01
13	51	68.0	582	3	O74931	ALTERNATIVE NADH-DEHYD	1.55e+01
14	51	68.0	1035	13	O57537	NF-PROTOCADHERIN.	1.55e+01
15	51	68.0	1069	11	O88185	BH-PROTOCADHERIN-A.	1.55e+01
16	51	68.0	1069	4	O60245	PCDH7 (BH-PCDH)A.	1.55e+01
17	51	68.0	1072	4	O60246	PCDH7 (BH-PCDH)B.	1.55e+01
18	51	68.0	1200	4	O60247	PCDH7 (BH-PCDH)C.	1.55e+01
19	50	66.7	293	2	O51095	CONSERVED HYPOTHETICAL	2.38e+01
20	50	66.7	444	5	P91141	SIMILAR TO ACETYLTRANS	2.38e+01
21	50	66.7	469	2	O32204	YVSH PROTEIN.	2.38e+01
22	50	66.7	626	2	Q46732	COLICIN PROTEIN.	2.38e+01
23	50	66.7	626	2	Q46734	COLICIN PROTEIN.	2.38e+01
24	50	66.7	626	2	Q46738	COLICIN PROTEIN.	2.38e+01
25	50	66.7	626	2	Q46740	COLICIN PROTEIN.	2.38e+01
26	50	66.7	989	11	Q63052	APOLIPOPROTEIN B (FRAG	2.38e+01
27	50	66.7	1186	14	O55767	PUTATIVE TYROSINE PROT	2.38e+01
28	49	65.3	112	3	Q07255	INTERNAL ORF OF L3149	3.65e+01
29	49	65.3	204	3	Q12097	HYPOTHETICAL 22.6 KD P	3.65e+01
30	49	65.3	222	13	P70023	OLFACTORY RECEPTOR (FR	3.65e+01
31	49	65.3	377	2	O84397	HYPOTHETICAL 41.4 KD P	3.65e+01
32	49	65.3	384	2	Q11197	HYPOTHETICAL 41.9 KD P	3.65e+01
33	49	65.3	392	2	O50873	CONSERVED HYPOTHETICAL	3.65e+01
34	49	65.3	409	2	O25853	NADH-UBIQUINONE OXIDOR	3.65e+01
35	49	65.3	650	2	O50417	MULTI-FUNCTIONAL ENZYM	3.65e+01
36	49	65.3	1026	4	Q08174	PROTOCADHERIN 42 PRECU	3.65e+01
37	49	65.3	1406	4	O15082	KIAA0377.	3.65e+01
38	48	64.0	170	2	O86052	AMMONIUM TRANSPORTER (	5.57e+01
39	48	64.0	325	2	Q55392	HYPOTHETICAL 35.4 KD P	5.57e+01
40	48	64.0	420	3	P78765	FISSION YEAST (FRAGMEN	5.57e+01
41	48	64.0	438	2	O67997	AMTB.	5.57e+01
42	48	64.0	505	11	O88808	TUBBY PROTEIN.	5.57e+01
43	48	64.0	954	2	O84594	HYPOTHETICAL 109.2 KD	5.57e+01
44	48	64.0	966	11	O55098	SERINE/THREONINE KINAS	5.57e+01
45	48	64.0	1685	10	O04142	RNA POLYMERASE II LARG	5.57e+01

ALIGNMENTS

RESULT 1
ID O60167 PRELIMINARY; PRT; 649 AA.
AC O60167;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PROTEIN COMPLEX ASSEMBLY PROTEIN.
GN SPBC19F8.03C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA BECK A., REINHARDT R., WOOD V., RAJANDREAM M.A., BARRELL B.G.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AL023594; E1293401; -.
SQ SEQUENCE 649 AA; 72985 MW; 9C207DB2 CRC32;

Query Match 73.3%; Score 55; DB 3; Length 649;
Best Local Similarity 54.5%; Pred. No. 2.58e+00;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 234 NYLSTARSLEF 244
:|||| : ||::
Qy 2 SYLSTSSSLDY 12

RESULT 2
ID O17959 PRELIMINARY; PRT; 349 AA.
AC O17959;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE M01B2.5 PROTEIN.
GN M01B2.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA LLOYD C.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.

```
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z83116; E1348127; -.
SQ   SEQUENCE   349 AA;  40017 MW;  90870FE2 CRC32;

  Query Match           72.0%;  Score 54;  DB 5;  Length 349;
  Best Local Similarity 70.0%;  Pred. No. 4.07e+00;
  Matches    7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    198 YLENSSGLDY 207
          ||  ||:|||
Qy      3 YLSTSSSLDY 12


RESULT    3
ID   O26074      PRELIMINARY;      PRT;   503 AA.
AC   O26074;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PROTEIN-EXPORT MEMBRANE PROTEIN (SECD).
GN   HP1550.
OS   HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC   BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC   HELICOBACTER.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695;
RX   MEDLINE; 97394467.
RA   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori[published erratum appears in Nature 1997 Sep
RT   25;389(6649):412].";
RL   NATURE 388:539-547(1997).
DR   EMBL; AE000652; G2314730; -.
DR   TIGR; HP1550; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   503 AA;  54247 MW;  9A76592C CRC32;

  Query Match           72.0%;  Score 54;  DB 2;  Length 503;
  Best Local Similarity 60.0%;  Pred. No. 4.07e+00;
  Matches    6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     38 YLSLASALEY 47
          ||| :|:|:|
Qy      3 YLSTSSSLDY 12


RESULT    4
ID   O11696      PRELIMINARY;      PRT;   127 AA.
AC   O11696;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NUCLEOPROTEIN (FRAGMENT).
OS   MEASLES VIRUS (SUBACUTE SCLEROSE PANENCEPHALITIS VIRUS).
OC   VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC   PARAMYXOVIRIDAE; PARAMYXOVIRINAE; MORBILLIVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=92-E;
RX   MEDLINE; 97278133.
RA   YAMAGUCHI S.;
RT   "Identification of three lineages of wild measles virus by nucleotide
RT   sequence analysis of N, P, M, F, and L genes in Japan.";
RL   J. MED. VIROL. 52:113-120(1997).
DR   EMBL; D87487; D1020995; -.
KW   NUCLEOPROTEIN.
FT   NON_TER       1      1
SQ   SEQUENCE   127 AA;  13950 MW;  42D75A2C CRC32;

  Query Match           70.7%;  Score 53;  DB 14;  Length 127;
  Best Local Similarity 63.6%;  Pred. No. 6.38e+00;
  Matches    7;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     66 AAYLPTSTPLD 76
          |:||:||::||
Qy      1 ASYLSTSSSLD 11


RESULT    5
ID   Q25144      PRELIMINARY;      PRT;   265 AA.
AC   Q25144;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HROX1.
GN   HROX1.
OS   HALIOTIS RUFESCENS (CALIFORNIA RED ABALONE).
OC   EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PROSOBRANCHIA;
OC   ARCHAEOGASTROPODA; HALIOTIDAE; HALIOTIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97388321.
RA   DEGNAN B.M., DEGNAN S.M., FENTENANY G., MORSE D.E.;
RT   "A Mox homeobox gene in the gastropod mollusc Haliotis rufescens is
RT   differentially expressed during larval morphogenesis and
RT   metamorphosis.";
RL   FEBS LETT. 411:119-122(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; X75217; G407415; -.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PFAM; PF00046; homeobox; 1.
KW   HOMEOBOX; DNA-BINDING; NUCLEAR PROTEIN.
SQ   SEQUENCE   265 AA;  29579 MW;  B68A753D CRC32;

  Query Match           70.7%;  Score 53;  DB 5;  Length 265;
  Best Local Similarity 75.0%;  Pred. No. 6.38e+00;
  Matches    9;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 GSYLSMSSSKDY 92
          :|||| ||| ||
Qy      1 ASYLSTSSSLDY 12


RESULT    6
ID   Q17336      PRELIMINARY;      PRT;   897 AA.
AC   Q17336;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   LET 858.
GN   LET-858.
OS   CAENORHABDITIS ELEGANS.
```

```
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2 (BRISTOL);
RA   KELLY W.G., COLES L.H., FIRE A.Z.;
RL   GENETICS 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MATTHEWS L.;
RL   SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U19615; G987227; -.
DR   EMBL; Z81525; E1351661; -.
SQ   SEQUENCE   897 AA;  104268 MW;  E1E3EA36 CRC32;

  Query Match           70.7%;  Score 53;  DB 5;  Length 897;
  Best Local Similarity  70.0%;  Pred. No. 6.38e+00;
  Matches    7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

      491 YLTMQSSLDY 500
          ||:  |||||
        3 YLSTSSSLDY 12


RESULT    7
ID   O61198      PRELIMINARY;      PRT;  1238 AA.
AC   O61198;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F15E6.6 PROTEIN.
GN   F15E6.6.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   MILLER N., STELLYES L., BRADSHAW H., KEPPLER D.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF038614; G2702437; -.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW   IRON-SULFUR.
SQ   SEQUENCE   1238 AA;  135726 MW;  9408BB7C CRC32;

  Query Match           70.7%;  Score 53;  DB 5;  Length 1238;
  Best Local Similarity  70.0%;  Pred. No. 6.38e+00;
  Matches    7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    702 YLETQSSLNY 711
          || | |||:|
Qy      3 YLSTSSSLDY 12


RESULT    8
ID   P91143      PRELIMINARY;      PRT;  372 AA.
AC   P91143;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO ACETYLTRANSFERASES.
GN   C37H5.2.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DAVIDSON S., GILLAM B.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U88315; G1825777; -.
DR   PFAM; PF00561; abhydrolase; 1.
KW   TRANSFERASE.
SQ   SEQUENCE   372 AA;  42139 MW;  5214F159 CRC32;

  Query Match           69.3%;  Score 52;  DB 5;  Length 372;
  Best Local Similarity  58.3%;  Pred. No. 9.96e+00;
  Matches    7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    168 GGYLSTSYALKY 179
          ::|||| :| |
Qy      1 ASYLSTSSSLDY 12


RESULT    9
ID   O93514      PRELIMINARY;      PRT;  475 AA.
AC   O93514;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   AXIAL PROTOCADHERIN (FRAGMENT).
GN   AXPC.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YAMAMOTO A., DEROBERTIS E.M.;
RT   "Xenopus axial protocadherin.";
RL   SUBMITTED (MAR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF053469; G3598688; -.
FT   NON_TER       1       1
```

```
FT   NON_TER     475    475
SQ   SEQUENCE   475 AA;  52268 MW;  2A681544 CRC32;

  Query Match           69.3%;  Score 52;  DB 13;  Length 475;
  Best Local Similarity  60.0%;  Pred. No. 9.96e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    341 FLQTTTSLDY 350
          :| |::|||| 
Qy      3 YLSTSSSLDY 12


RESULT   10
ID   O01623      PRELIMINARY;      PRT;   932 AA.
AC   O01623;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO LIGAND-GATED IONIC CHANNEL PROTEINS.
     ZC196.7.
     CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   MURRAY J.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
     WATERSTON R.;
     SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U97007; G1938466; -.
DR   PFAM; PF00060; lig_chan; 2.
SQ   SEQUENCE   932 AA;  106836 MW;  469F21FA CRC32;

  Query Match           69.3%;  Score 52;  DB 5;  Length 932;
  Best Local Similarity  54.5%;  Pred. No. 9.96e+00;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    753 AYLMESTSLEY 763
          :||  |:||:|
Qy      2 SYLSTSSSLDY 12


RESULT   11
ID   O48385      PRELIMINARY;      PRT;    83 AA.
AC   O48385;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ORF83.
OS   STREPTOCOCCUS THERMOPHILUS BACTERIOPHAGE TP-J34.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TP-J34;
RX   MEDLINE; 98122991.
RA   NEVE H., ZENZ K.I., DESIERE F., KOCH A., HELLER K.J., BRUSSOW H.;
RT   "Comparison of the lysogeny modules from the temperate Streptococcus
RT   thermophilus bacteriophages TP-J34 and Sfi21: implications for the
RT   modular theory of phage evolution.";
RL   VIROLOGY 241:61-72(1998).
DR   EMBL; AF020798; G2897100; -.
SQ   SEQUENCE   83 AA;  9876 MW;  250B3F9C CRC32;

  Query Match           68.0%;  Score 51;  DB 9;  Length 83;
  Best Local Similarity  60.0%;  Pred. No. 1.55e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db     71 YLETSAFLEY 80
          || ||: |:|
Qy      3 YLSTSSSLDY 12


RESULT   12
ID   O23552      PRELIMINARY;      PRT;   471 AA.
AC   O23552;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 52.8 KD PROTEIN.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BEVAN M., STIEKEMA W., MURPHY G., WAMBUTT R., POHL T., TERRYN N.,
RA   KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
RA   PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,
RA   JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
RA   SCHUELLER C., CHALWATZIS N.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z97342; E327038; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   471 AA;  52785 MW;  4EBA8315 CRC32;

  Query Match           68.0%;  Score 51;  DB 10;  Length 471;
  Best Local Similarity  63.6%;  Pred. No. 1.55e+01;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     19 AGYLTSSSDLD 29
          |:||::|| ||
Qy      1 ASYLSTSSSLD 11


RESULT   13
ID   O74931      PRELIMINARY;      PRT;   582 AA.
AC   O74931;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ALTERNATIVE NADH-DEHYDROGENASE PRECURSOR (EC 1.6.5.3)
DE   (NADH DEHYDROGENASE (UBIQUINONE)) (UBIQUINONE REDUCTASE)
DE   (TYPE I DEHYDROGENASE) (COMPLEX I DEHYDROGENASE).
GN   NDH2.
OS   YARROWIA LIPOLYTICA (CANDIDA LIPOLYTICA).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   DIPODASCACEAE; YARROWIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=E150;
RA   KERSCHER S.J., BRANDT U.;
```

```
RT   "Identification of the YLNDH2 Gene Encoding the Alternative
RT   NADH:Ubiquinone Oxidoreductase from Yarrowia lipolytica.";
RL   SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   -!- COFACTOR: FAD; IRON-SULFUR.
DR   EMBL; AJ006852; E1330342; -.
KW   SIGNAL; OXIDOREDUCTASE.
FT   SIGNAL        1     79       POTENTIAL.
FT   CHAIN        80    582       ALTERNATIVE NADH-DEHYDROGENASE.
SQ   SEQUENCE   582 AA;  65814 MW;  0460C796 CRC32;

  Query Match           68.0%;  Score 51;  DB 3;  Length 582;
  Best Local Similarity  54.5%;  Pred. No. 1.55e+01;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 TYLSAIASLGY 91
          :|||: :|| |
Qy      2 SYLSTSSSLDY 12


ULT   14
    O57537      PRELIMINARY;      PRT;  1035 AA.
AC   O57537;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE   NF-PROTOCADHERIN.
GN   NFPC.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BRADLEY R.S., ESPESETH A., KINTNER C.;
RL   CURR. BIOL. 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF043643; G2852363; -.
DR   PROSITE; PS00232; CADHERIN; 6.
KW   CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ   SEQUENCE   1035 AA;  113713 MW;  7E4D3C4E CRC32;

  Query Match           68.0%;  Score 51;  DB 13;  Length 1035;
  Best Local Similarity  60.0%;  Pred. No. 1.55e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    458 FLHTSAPLDY 467
          :| ||::|||
Qy      3 YLSTSSSLDY 12


ESULT   15
ID   O88185      PRELIMINARY;      PRT;  1069 AA.
AC   O88185;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   BH-PROTOCADHERIN-A.
GN   PCDH7.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98277460.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RT   "Cloning, expression analysis, and chromosomal localization of
RT   BH-protocadherin (PCDH7), a novel member of the cadherin
RT   superfamily.";
RL   GENOMICS 49:458-461(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., HIDA M., WATANABE M., YAMAGUCHI R., TATEYAMA S.,
RA   SUGANO S.;
RT   "cDNA cloning and chromosomal mapping of mouse BH-protocadherin.";
RL   DNA SEQ. 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006758; D1033562; -.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ   SEQUENCE   1069 AA;  116314 MW;  0F3F60C6 CRC32;

  Query Match           68.0%;  Score 51;  DB 11;  Length 1069;
  Best Local Similarity  60.0%;  Pred. No. 1.55e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    486 FLHTSAPLDY 495
          :| ||::|||
Qy      3 YLSTSSSLDY 12


Search completed: Thu Sep  2 12:32:56 1999
Job time : 50 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:34:45 1999; MasPar time 3.53 Seconds
72.272 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-30
Description: (1-12) from US08599226.pep
Perfect Score: 69
Sequence: 1 ASYLSTSSSLDD 12

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 16.937; Variance 50.990; scale 0.332

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	69	100.0	12	27	W27589	Anti-TNF-alpha antibo	7.15e-01
2	66	95.7	12	27	W27586	Anti-TNF-alpha antibo	1.59e+00
3	63	91.3	12	27	W27588	Anti-TNF-alpha antibo	3.52e+00
4	63	91.3	12	27	W27587	Anti-TNF-alpha antibo	3.52e+00
5	60	87.0	12	27	W27593	Anti-TNF-alpha antibo	7.70e+00
6	60	87.0	12	27	W27594	Anti-TNF-alpha antibo	7.70e+00
7	57	82.6	12	27	W27563	Anti-TNF-alpha antibo	1.67e+01
8	57	82.6	12	27	W27592	Anti-TNF-alpha antibo	1.67e+01
9	57	82.6	12	27	W27591	Anti-TNF-alpha antibo	1.67e+01
10	57	82.6	121	27	W27569	Anti-TNF-alpha antibo	1.67e+01
11	55	79.7	12	27	W27590	Anti-TNF-alpha antibo	2.78e+01
12	50	72.5	417	26	W23067	Canine IgE heavy chai	9.68e+01
13	49	71.0	314	15	R87027	T7 gene 10 leader seq	1.24e+02
14	49	71.0	384	20	W06968	M. tuberculosis ORF 2	1.24e+02
15	49	71.0	386	36	W68459	PRRSV isolate 14/96 n	1.24e+02
16	49	71.0	393	36	W68458	PRRSV isolate 4/96 nu	1.24e+02
17	48	69.6	2799	39	W81867	Human tumour suppress	1.58e+02
18	47	68.1	212	33	W62322	Pseudomonas fluoresce	2.01e+02
19	47	68.1	213	28	W33747	Pseudomonas gafA elem	2.01e+02
20	47	68.1	213	32	W37952	Amino acid sequence o	2.01e+02
21	47	68.1	213	27	W40804	ORF5 protein (gafA).	2.01e+02
22	47	68.1	213	25	W30202	ORF 5 protein (gafA).	2.01e+02
23	47	68.1	213	9	R45540	Polypeptide encoded b	2.01e+02
24	47	68.1	213	27	W38359	Protein encoded by OR	2.01e+02
25	47	68.1	213	25	W30206	Ile 182 ORF 5 protein	2.01e+02
26	47	68.1	213	33	W62321	Pseudomonas fluoresce	2.01e+02
27	47	68.1	213	32	W37951	Amino acid sequence o	2.01e+02
28	47	68.1	213	33	W62320	Pseudomonas fluoresce	2.01e+02
29	46	66.7	19	19	R95335	MBP-1.1 (11-29).	2.55e+02
30	46	66.7	20	39	W73609	Human myelin basic pr	2.55e+02
31	46	66.7	20	27	W37550	Human myelin basic pr	2.55e+02
32	46	66.7	21	27	W37552	Human myelin basic pr	2.55e+02
33	46	66.7	21	19	R95336	MBP-1.2 (11-31).	2.55e+02
34	46	66.7	127	9	R48596	Rat myelin basic prot	2.55e+02
35	46	66.7	167	9	R48595	Myelin basic protein.	2.55e+02
36	46	66.7	168	9	R48594	Rabbit myelin basic p	2.55e+02
37	46	66.7	170	6	R30736	Human MBP.	2.55e+02
38	46	66.7	170	9	R48592	Human myelin basic pr	2.55e+02
39	46	66.7	171	18	R99580	Human myelin basic pr	2.55e+02
40	46	66.7	311	38	W89812	Protein encoded by cl	2.55e+02
41	46	66.7	373	19	W06103	MP4 chimera (MBP21.5-	2.55e+02
42	46	66.7	481	30	W40054	P300/CBP-associated t	2.55e+02
43	46	66.7	832	30	W40052	Human P300/CBP-associ	2.55e+02
44	46	66.7	1091	27	W41641	Sequence used in dete	2.55e+02
45	46	66.7	1146	39	W86007	Histidine kinase CgSl	2.55e+02

ALIGNMENTS

RESULT 1
ID W27589 standard; peptide; 12 AA.
AC W27589;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          100.0%;  Score 69;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 7.15e-01;
  Matches   12;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssldd 12
          ||||||||||||
Qy      1 ASYLSTSSSLDD 12


RESULT    2
ID   W27586 standard; peptide; 12 AA.
AC   W27586;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 72; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          95.7%;  Score 66;  DB 27;  Length 12;
  Best Local Similarity 91.7%;  Pred. No. 1.59e+00;
  Matches   11;  Conservative   1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssldn 12
          ||||||||||| :
Qy      1 ASYLSTSSSLDD 12


RESULT    3
ID   W27588 standard; peptide; 12 AA.
AC   W27588;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          91.3%;  Score 63;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 3.52e+00;
  Matches   11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssld 11
          ||||||||||
Qy      1 ASYLSTSSSLD 11


RESULT    4
ID   W27587 standard; peptide; 12 AA.
AC   W27587;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
```

```
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           91.3%;  Score 63;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 3.52e+00;
  Matches    11;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 asylstsssld 11
           ||||||||||:
Qy       1 ASYLSTSSSLD 11


RESULT    5
ID   W27593 standard; peptide; 12 AA.
AC   W27593;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 75; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           87.0%;  Score 60;  DB 27;  Length 12;
  Best Local Similarity 81.8%;  Pred. No. 7.70e+00;
  Matches     9;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 asylstassle 11
           ||||||:|||:
Qy       1 ASYLSTSSSLD 11


RESULT    6
ID   W27594 standard; peptide; 12 AA.
AC   W27594;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Disclosure; Page 75; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
```

```
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           87.0%;  Score 60;  DB 27;  Length 12;
  Best Local Similarity  81.8%;  Pred. No. 7.70e+00;
  Matches    9;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 sylstassldn 12
          |||||:||||:
Qy      2 SYLSTSSSLDD 12


RESULT    7
ID   W27563 standard; peptide; 12 AA.
AC   W27563;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 12
FT                   /label= Tyr, Asn
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 9; Page 65; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           82.6%;  Score 57;  DB 27;  Length 12;
  Best Local Similarity  90.0%;  Pred. No. 1.67e+01;
  Matches    9;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 sylstassld 11
          |||||:||||
Qy      2 SYLSTSSSLD 11


RESULT    8
ID   W27592 standard; peptide; 12 AA.
AC   W27592;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           82.6%;  Score 57;  DB 27;  Length 12;
  Best Local Similarity  81.8%;  Pred. No. 1.67e+01;
  Matches    9;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db      1 asflstsssle 11
            ||:|||||||:
Qy      1 ASYLSTSSSLD 11


RESULT    9
ID   W27591 standard; peptide; 12 AA.
AC   W27591;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           82.6%;  Score 57;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 1.67e+01;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssl 10
            ||||||||||
Qy      1 ASYLSTSSSL 10


RESULT   10
ID   W27569 standard; Protein; 121 AA.
AC   W27569;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain variable region.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody;
KW   heavy chain; variable region; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
DR   N-PSDB; T88404.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 16; Page 76; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain variable region.
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   121 AA;

  Query Match           82.6%;  Score 57;  DB 27;  Length 121;
  Best Local Similarity 90.0%;  Pred. No. 1.67e+01;
  Matches     9;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    100 sylstassld 109
            |||||:||||
Qy      2 SYLSTSSSLD 11


RESULT   11
ID   W27590 standard; peptide; 12 AA.
AC   W27590;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
```

```
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           79.7%;  Score 55;  DB 27;  Length 12;
  Best Local Similarity  90.9%;  Pred. No. 2.78e+01;
  Matches    10;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 asylstsfsld 11
          ||||||| |||
Qy      1 ASYLSTSSSLD 11


RESULT   12
ID   W23067 standard; Protein; 417 AA.
AC   W23067;
DT   19-FEB-1998  (first entry)
DE   Canine IgE heavy chain constant region (exon 1-4 product).
KW   IgE; immunoglobulin; antibody; heavy chain constant region;
KW   allergy; hypersensitivity; therapy; dog; antisense;
KW   immunomodulation.
OS   Canis familiaris.
FH   Key             Location/Qualifiers
FT   Misc_difference 55
FT                   /note= "encoded by ACC"
FT   Misc_difference 56
FT                   /note= "encoded by TAC"
FT   Misc_difference 67
FT                   /note= "encoded by GCC"
FT   Misc_difference 83
FT                   /note= "encoded by NNT"
FT   Misc_difference 174
FT                   /note= "encoded by GGN"
FT   Misc_difference 175
FT                   /note= "encoded by NNG"
FT   Misc_difference 176
FT                   /note= "encoded by TGN"
FT   Misc_difference 203
FT                   /note= "encoded by TCC"
FT   Misc_difference 204
FT                   /note= "encoded by GAC"
PN   WO9730156-A2.
PD   21-AUG-1997.
PF   14-FEB-1997; U02322.
PR   14-FEB-1996; US-601197.
PA   (IDEX-) IDEXX LAB INC.
PI   Harris RA,  Mermer B,  Siefring AE;
DR   WPI; 97-425031/39.
DR   N-PSDB; T79278.
PT   Isolated canine IgE heavy chain constant region DNA - useful to
PT   develop products for treatment of canine allergies and for
PT   immunomodulation in dogs
PS   Disclosure; Page 35-39; 59pp; English.
CC   This polypeptide is encoded by exons 1-4 (see T79278) of canine
CC   IgE heavy chain constant region (epsilon) genomic DNA.  Another
CC   polypeptide, comprising the exon 5 and 6 product, is given in
CC   W23068.  Recombinant peptides encoded by exons 1-6 can be
CC   produced in eukaryotic or prokaryotic cells.  Such peptides,
CC   and antibodies raised against them, are used in methods to treat
CC   the manifestation of allergy in dogs, e.g. to treatment Type I
CC   immediate hypersensitivity, and for immunomodulation.
SQ   Sequence   417 AA;

  Query Match           72.5%;  Score 50;  DB 26;  Length 417;
  Best Local Similarity  63.6%;  Pred. No. 9.68e+01;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    209 ssylsppspld 219
          :|||| :|:||
Qy      1 ASYLSTSSSLD 11


RESULT   13
ID   R87027 standard; Protein; 314 AA.
AC   R87027;
DT   18-MAR-1996  (first entry)
DE   T7 gene 10 leader sequence product.
KW   Plasmid pTOPE-1b(+); vector; ligand binding domain; epitope mapping;
KW   antigen; Escherichia coli.
OS   Bacteriophage T7.
PN   US5464745-A.
PD   07-NOV-1995.
PF   31-MAR-1993; 040753.
PR   31-MAR-1993; US-040753.
PA   (NOVA-) NOVAGEN INC.
PI   Garber R,  Hammer B,  Mierendorf R,  Novy R;
DR   WPI; 95-392610/50.
DR   N-PSDB; T07310.
PT   Mapping ligand binding domains, esp. epitope(s), of proteins - by
PT   expressing peptide(s) encoded by random gene fragments and testing
PT   for ligand binding
PS   Disclosure; Column 15-18; 12pp; English.
CC   The phage T7 gene 10 leader sequence product (R87027) is encoded
CC   by prokaryotic expression vector pTOPE-1b(+) (T07310).  A fusion
CC   of the leader sequence and a putative ligand-binding domain (LBD),
CC   esp. antigenic region, of a protein is obtd. following insertion
CC   of putative LBD-encoding DNA into the vector.  The fusion protein
CC   accumulates as inclusion bodies in Escherichia coli host cells and
CC   can be screened for its ability to bind a ligand.
SQ   Sequence   314 AA;

  Query Match           71.0%;  Score 49;  DB 15;  Length 314;
  Best Local Similarity  50.0%;  Pred. No. 1.24e+02;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     71 aaylapgenldd 82
          |:||: : :|||
Qy      1 ASYLSTSSSLDD 12


RESULT   14
ID   W06968 standard; Protein; 384 AA.
AC   W06968;
DT   12-MAR-1997  (first entry)
DE   M. tuberculosis ORF 2 product.
KW   MACE; mycolic acid cyclopropanating enzyme; cyclopropanation; cma;
KW   pathogenic form; mycobacterium; non-pathogenic; screen; inhibitor;
KW   cyclopropane mycolic acid synthase; beta-ketoacyl reductase;
KW   Streptomyces cinnamonesis; homology.
OS   Mycobacterium tuberculosis.
```

PN US5573915-A.
PD 12-NOV-1996.
PF 01-JUN-1995; 457245.
PR 01-JUN-1995; US-457245.
PA (USGO) US GOVERNMENT.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Clifton B, Yuan Y;
DR WPI; 96-517880/51.
DR N-PSDB; T46163.
PT Screening assay for anti-mycobacterial agents - based on inhibition
PT of mycolic acid cyclo:propanating enzyme
PS Disclosure; Column 35-38; 35pp; English.
CC This sequence is encoded by ORF 2 of Mycobacterium tuberculosis (see
CC T46163). Three ORFs have been found all within a 7.2 kb BamHI fragment
CC isolated from M. tuberculosis, and are believed to be related to the
CC biosynthesis of mycolic acids. The ORF 2 gene product has homology to
CC known enzymes involved in the oxidative/reductive interconversions of
CC a ketone and an alcohol. It is most homologous (30 percent identity
CC over 188 amino acids) to actIII, beta-ketoacyl reductase from
CC Streptomyces cinnamonesis which is involved in chain elongation in
CC polyketide biosynthesis. Cycloproponation of mycolic acids distinguishes
CC pathogenic forms of mycobacterium from non-pathogenic forms. A method to
CC determine the ability of a cpd. to inhibit cycloproponation of mycolic
CC acids in pathogenic mycobacterium is claimed.
SQ Sequence 384 AA;

Query Match 71.0%; Score 49; DB 20; Length 384;
Best Local Similarity 54.5%; Pred. No. 1.24e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 235 ssylptkaald 245
:|||:| ::||
Qy 1 ASYLSTSSSLD 11

RESULT 15
ID W68459 standard; protein; 386 AA.
AC W68459;
DT 06-JAN-1999 (first entry)
DE PRRSV isolate 14/96 nucleocapsid/phage T7 gene 10 fusion protein.
KW Nucleocapsid; Porcine Reproductive and Respiratory Syndróme Virus; PRRSV;
KW pig; serum; RT-PCR; reverse transcription; amplification; fusion protein;
KW primer; bacteriophage T7; gene 10; E coli; immunoassay; diagnosis.
OS Chimeric - Porcine reproductive and respiratory syndrome virus.
OS Chimeric - Bacteriophage T7.
PN WO9829550-A1.
PD 09-JUL-1998.
PF 24-DEC-1997; ES0313.
PR 30-DEC-1996; ES-002770.
PA (INMU-) INMUNOLOGIA & GENETICA APLICADA SA.
PI Casal Alvarez JI, Rodriguez Garcia MJ, Sanz Fernandez A;
DR WPI; 98-388131/33.
PT New fusion proteins comprising Porcine Reproductive and Respiratory
PT Syndrome Virus nucleocapsid - useful as reagents in immuno:diagnosis
PT of PRRS, are produced recombinantly in bacterial hosts
PS Claim 5; Page 31-32; 42pp; Spanish.
CC This sequence represents a fusion protein comprising the nucleocapsid
CC protein from the Porcine Reproductive and Respiratory Syndrome Virus
CC (PRRSV) American isolate Canada 14/96 fused to the gene 10 protein from
CC bacteriophage T7. The fusion protein is generated by cloning the PRRSV
CC nucleocapsid (W68457) coding sequence into the bacterial expression
CC vector pET3x. The fusion protein is produced when the gene is expressed
CC in E. coli (BL21) pLysS cells. The nucleocapsid protein, and especially
CC the fusion protein, is useful as a reagent in immunoassays to diagnose
CC PRRSV.
SQ Sequence 386 AA;

Query Match 71.0%; Score 49; DB 36; Length 386;
Best Local Similarity 50.0%; Pred. No. 1.24e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 71 aaylapgenldd 82
|:||: : :|||
Qy 1 ASYLSTSSSLDD 12

Search completed: Thu Sep 2 12:35:05 1999
Job time : 20 secs.

****

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:37:02 1999; MasPar time 1.38 Seconds
88.598 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-30
Description: (1-12) from US08599226.pep
Perfect Score: 69
Sequence: 1 ASYLSTSSSLDD 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 15.800; Variance 47.088; scale 0.336

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	50	72.5	426	3	PCT-US95-1	Sequence 2, Applicatio	3.45e+01
2	50	72.5	426	1	US-08-336-	Sequence 2, Applicatio	3.45e+01
3	49	71.0	314	1	US-08-040-	Sequence 2, Applicatio	4.46e+01
4	49	71.0	384	1	US-08-457-	Sequence 5, Applicatio	4.46e+01
5	47	68.1	213	2	US-08-761-	Sequence 10, Applicati	7.38e+01
6	47	68.1	213	1	US-08-287-	Sequence 2, Applicatio	7.38e+01
7	47	68.1	213	2	US-08-761-	Sequence 5, Applicatio	7.38e+01
8	47	68.1	213	1	US-08-459-	Sequence 2, Applicatio	7.38e+01
9	47	68.1	213	1	US-08-459-	Sequence 2, Applicatio	7.38e+01
10	47	68.1	213	1	US-08-460-	Sequence 2, Applicatio	7.38e+01
11	47	68.1	213	2	US-08-761-	Sequence 3, Applicatio	7.38e+01
12	47	68.1	213	3	PCT-US93-0	Sequence 2, Applicatio	7.38e+01
13	46	66.7	20	2	US-08-468-	Sequence 2, Applicatio	9.47e+01
14	46	66.7	21	2	US-08-787-	Sequence 33, Applicati	9.47e+01
15	46	66.7	170	2	US-08-327-	Sequence 1, Applicatio	9.47e+01
16	45	65.2	168	4	5194425-4	Patent No. 5194425.	1.21e+02
17	45	65.2	432	2	US-08-705-	Sequence 18, Applicati	1.21e+02
18	44	63.8	209	2	US-08-771-	Sequence 2, Applicatio	1.55e+02
19	44	63.8	464	1	US-07-688-	Sequence 16, Applicati	1.55e+02
20	44	63.8	464	3	PCT-US91-0	Sequence 16, Applicati	1.55e+02
21	43	62.3	170	4	5194425-3	Patent No. 5194425.	1.98e+02
22	43	62.3	170	2	US-08-227-	Sequence 1, Applicatio	1.98e+02
23	43	62.3	238	2	US-08-928-	Sequence 5, Applicatio	1.98e+02
24	43	62.3	240	3	PCT-US95-0	Sequence 80, Applicati	1.98e+02
25	43	62.3	240	1	US-08-261-	Sequence 80, Applicati	1.98e+02
26	43	62.3	345	2	US-08-446-	Sequence 40, Applicati	1.98e+02
27	43	62.3	1255	2	US-08-468-	Sequence 68, Applicati	1.98e+02
28	43	62.3	1255	2	US-08-484-	Sequence 8, Applicatio	1.98e+02
29	43	62.3	1255	1	US-08-467-	Sequence 68, Applicati	1.98e+02
30	43	62.3	1255	2	US-08-486-	Sequence 68, Applicati	1.98e+02
31	43	62.3	1255	2	US-08-625-	Sequence 2, Applicatio	1.98e+02
32	43	62.3	1255	2	US-08-356-	Sequence 2, Applicatio	1.98e+02
33	43	62.3	1255	2	US-08-414-	Sequence 68, Applicati	1.98e+02
34	42	60.9	158	2	US-08-828-	Sequence 3, Applicatio	2.53e+02
35	42	60.9	159	2	US-08-828-	Sequence 1, Applicatio	2.53e+02
36	42	60.9	326	4	5395759-2	Patent No. 5395759.	2.53e+02
37	42	60.9	822	2	US-08-222-	Sequence 7, Applicatio	2.53e+02
38	42	60.9	3666	2	US-08-222-	Sequence 12, Applicati	2.53e+02
39	42	60.9	3727	2	US-08-222-	Sequence 27, Applicati	2.53e+02
40	42	60.9	3778	2	US-08-222-	Sequence 2, Applicatio	2.53e+02
41	41	59.4	113	1	US-08-241-	Sequence 10, Applicati	3.21e+02
42	41	59.4	536	1	US-08-401-	Sequence 2, Applicatio	3.21e+02
43	41	59.4	553	3	PCT-US94-0	Sequence 4, Applicatio	3.21e+02
44	41	59.4	554	2	US-08-464-	Sequence 1, Applicatio	3.21e+02
45	41	59.4	879	2	US-08-486-	Sequence 6, Applicatio	3.21e+02

ALIGNMENTS

```
RESULT    1
ID   PCT-US95-13795-2     STANDARD;     PRT;   426 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9513795
XX
CC    Sequence 2, Application PC/TUS9513795
CC     GENERAL INFORMATION:
CC       APPLICANT:  HOLLIS, GREGORY F.
CC       APPLICANT:  PATEL, MAYUR D.
CC       TITLE OF INVENTION:  DNA ENCODING CANINE IMMUNOGLOBULINS
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  126 E. LINCOLN AVENUE; P.O. BOX 2000
CC         CITY:  RAHWAY
CC         STATE:  NEW JERSEY
CC         COUNTRY:  USA
CC         ZIP:  07065-0907
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/13795
CC         FILING DATE:
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  19211Y
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (908) 594-6734
CC         TELEFAX:  (908) 594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  426 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   426 AA; 47234 MW; 1032622 CN;
```

```
  Query Match           72.5%;  Score 50;  DB 3;  Length 426;
  Best Local Similarity  63.6%;  Pred. No. 3.45e+01;
  Matches    7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    214 TSYLSPPSPLD 224
          :|||| :|:||
Qy      1 ASYLSTSSSLD 11


RESULT    2
ID   US-08-336-583-2     STANDARD;     PRT;   426 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08336583
XX
CC    Sequence 2, Application US/08336583
CC    Patent No. 5629415
CC     GENERAL INFORMATION:
CC       APPLICANT:  HOLLIS, GREGORY F.
CC       APPLICANT:  PATEL, MAYUR D.
CC       TITLE OF INVENTION:  DNA ENCODING CANINE IMMUNOGLOBULIN E
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  126 E. LINCOLN AVENUE
CC         CITY:  RAHWAY
CC         STATE:  NEW JERSEY
CC         COUNTRY:  USA
CC         ZIP:  07065-0900
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/336,583
CC         FILING DATE:  09-NOV-1994
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  19211
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (908) 594-6734
CC         TELEFAX:  (908) 594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  426 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   426 AA; 47234 MW; 1032622 CN;

  Query Match           72.5%;  Score 50;  DB 1;  Length 426;
  Best Local Similarity  63.6%;  Pred. No. 3.45e+01;
  Matches    7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    214 TSYLSPPSPLD 224
          :|||| :|:||
Qy      1 ASYLSTSSSLD 11


RESULT    3
ID   US-08-040-753-2     STANDARD;     PRT;   314 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08040753
XX
CC    Sequence 2, Application US/08040753
CC    Patent No. 5464745
CC     GENERAL INFORMATION:
CC       APPLICANT:  Mierendorf, Robert
CC       APPLICANT:  Garber, Richard
CC       APPLICANT:  No. 5464745y, Robert
CC       APPLICANT:  Hammer, Beth
CC       TITLE OF INVENTION:  Protein Ligand Binding
CC       TITLE OF INVENTION:  Region Mapping System
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Quarles and Brady
CC         STREET:  1 South Pinckney St., Box 2113
CC         CITY:  Madison
CC         STATE:  WI
CC         COUNTRY:  USA
CC         ZIP:  53701-2113
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/040,753
CC         FILING DATE:  19930331
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Seay, Nicholas J
CC         REGISTRATION NUMBER:  27386
CC         REFERENCE/DOCKET NUMBER:  70-399-9001-1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  608-251-5000
CC         TELEFAX:  608-251-9166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  314 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   314 AA; 33533 MW; 495848 CN;

  Query Match           71.0%;  Score 49;  DB 1;  Length 314;
  Best Local Similarity  50.0%;  Pred. No. 4.46e+01;
  Matches    6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     71 AAYLAPGENLDD 82
          |:||: : :|||
Qy      1 ASYLSTSSSLDD 12


RESULT    4
ID   US-08-457-245-5     STANDARD;     PRT;   384 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/08457245
XX
CC    Sequence 5, Application US/08457245
CC    Patent No. 5573915
CC     GENERAL INFORMATION:
CC       APPLICANT:  BARRY III, Clifton E.
CC       APPLICANT:  YUAN, Ying
CC       TITLE OF INVENTION:  CLONING AND EXPRESSION OF DNA INVOLVED
CC       TITLE OF INVENTION:  IN THE BIOSYNTHESIS OF CYCLOPROPANATED MYCOLIC ACIDS IN
CC       TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS
CC       NUMBER OF SEQUENCES:  21
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Townsend and Townsend Khourie and Crew
```

```
CC         STREET:  Steuart Street Tower, One Market Plaza
CC         CITY:  San Francisco
CC         STATE:  California
CC         COUNTRY:  US
CC         ZIP:  94105-1493
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/457,245
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Chambers, Guy W.
CC         REGISTRATION NUMBER:  30,617
CC         REFERENCE/DOCKET NUMBER:  15280-216000
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 543-9600
CC         TELEFAX:  (415) 543-5043
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  384 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  ORF2 protein
SQ   SEQUENCE   384 AA; 41963 MW; 701271 CN;

  Query Match           71.0%;  Score 49;  DB 1;  Length 384;
  Best Local Similarity  54.5%;  Pred. No. 4.46e+01;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    235 SSYLPTKAALD 245
          :|||:| ::||
Qy      1 ASYLSTSSSLD 11


RESULT    5
ID   US-08-761-258-10     STANDARD;     PRT;   213 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 10, Application US/08761258
XX
CC    Sequence 10, Application US/08761258
CC    Patent No. 5756087
CC     GENERAL INFORMATION:
CC       APPLICANT:  Ligon, James M.
CC       APPLICANT:  Hill, Dwight S.
CC       APPLICANT:  Lam, Stephen T.
CC       APPLICANT:  Gaffney, Thomas D.
CC       APPLICANT:  Torkewitz, Nancy
CC       TITLE OF INVENTION:  Genetically Modified Pseudomonas Strains
CC       TITLE OF INVENTION:  with Enhanced Biocontrol Activity
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Ciba-Geigy Corporation
CC         STREET:  520 White Plains Road, P.O. Box 2005
CC         CITY:  Tarrytown
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10591
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/761,258
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Meigs, J. Timothy
CC         REGISTRATION NUMBER:  38,241
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (919) 541-8587
CC         TELEFAX:  (919) 541-8689
CC     INFORMATION FOR SEQ ID NO:  10:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  213 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   213 AA; 23392 MW; 236428 CN;

  Query Match           68.1%;  Score 47;  DB 2;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 7.38e+01;
  Matches     4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db     99 AGYLTKGAGLNE 110
          |:||: :::|::
Qy      1 ASYLSTSSSLDD 12


RESULT    6
ID   US-08-287-442-2     STANDARD;     PRT;   213 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08287442
XX
CC    Sequence 2, Application US/08287442
CC    Patent No. 5670350
CC     GENERAL INFORMATION:
CC       APPLICANT:  Gaffney, Thomas D.
CC       APPLICANT:  Lam, Stephen T.
CC       APPLICANT:  Ligon, James M.
CC       APPLICANT:  Hill, Dwight S.
CC       APPLICANT:  Stein, Jeffrey I.
CC       APPLICANT:  Howell, Charles R.
CC       APPLICANT:  Becker, J. Ole
CC       TITLE OF INVENTION:  Gene Activating Element
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CIBA-GEIGY Corporation
CC         STREET:  7 Skyline Drive
CC         CITY:  Hawthorne
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10532
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/287,442
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/087,636
CC         FILING DATE:  01-JUL-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/908,284
CC         FILING DATE:  02-JUL-1992
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/570,184
CC         FILING DATE:  08-AUG-1990
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Elmer, James Scott
CC         REGISTRATION NUMBER:  36,129
```

```
CC         REFERENCE/DOCKET NUMBER:  CGC 1506/CIP4
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  919-541-8614
CC         TELEFAX:  919-541-8689
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  213 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   213 AA; 23364 MW; 236956 CN;

  Query Match           68.1%;  Score 47;  DB 1;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 7.38e+01;
  Matches    4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db     99 AGYLTKGAGLNE 110
          |:||: :::|::
Qy      1 ASYLSTSSSLDD 12


RESULT    7
ID   US-08-761-258-5     STANDARD;     PRT;   213 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/08761258
XX
CC    Sequence 5, Application US/08761258
CC    Patent No. 5756087
CC     GENERAL INFORMATION:
CC       APPLICANT:  Ligon, James M.
CC       APPLICANT:  Hill, Dwight S.
CC       APPLICANT:  Lam, Stephen T.
CC       APPLICANT:  Gaffney, Thomas D.
CC       APPLICANT:  Torkewitz, Nancy
CC       TITLE OF INVENTION:  Genetically Modified Pseudomonas Strains
CC       TITLE OF INVENTION:  with Enhanced Biocontrol Activity
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Ciba-Geigy Corporation
CC         STREET:  520 White Plains Road, P.O. Box 2005
CC         CITY:  Tarrytown
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10591
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/761,258
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Meigs, J. Timothy
CC         REGISTRATION NUMBER:  38,241
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (919) 541-8587
CC         TELEFAX:  (919) 541-8689
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  213 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   213 AA; 23364 MW; 236956 CN;

  Query Match           68.1%;  Score 47;  DB 2;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 7.38e+01;
  Matches    4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db     99 AGYLTKGAGLNE 110
          |:||: :::|::
Qy      1 ASYLSTSSSLDD 12


RESULT    8
ID   US-08-459-701-2     STANDARD;     PRT;   213 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08459701
XX
CC    Sequence 2, Application US/08459701
CC    Patent No. 5686282
CC     GENERAL INFORMATION:
CC       APPLICANT:  Gaffney, Thomas D.
CC       APPLICANT:  Lam, Stephen T.
CC       APPLICANT:  Ligon, James M.
CC       APPLICANT:  Hill, Dwight S.
CC       APPLICANT:  Stein, Jeffrey I.
CC       APPLICANT:  Howell, Charles R.
CC       APPLICANT:  Becker, J. Ole
CC       TITLE OF INVENTION:  Gene Activating Element
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CIBA-GEIGY Corporation
CC         STREET:  7 Skyline Drive
CC         CITY:  Hawthorne
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10532
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/459,701
CC         FILING DATE:  02-JUN-1995
CC         CLASSIFICATION:  536
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/287,442
CC         FILING DATE:  08-AUG-1994
CC         APPLICATION NUMBER:  US 08/087,636
CC         FILING DATE:  01-JUL-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/908,284
CC         FILING DATE:  02-JUL-1992
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/570,184
CC         FILING DATE:  08-AUG-1990
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Elmer, James Scott
CC         REGISTRATION NUMBER:  36,129
CC         REFERENCE/DOCKET NUMBER:  CGC 1506/CIP4
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  919-541-8614
CC         TELEFAX:  919-541-8689
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  213 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   213 AA; 23364 MW; 236956 CN;

  Query Match           68.1%;  Score 47;  DB 1;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 7.38e+01;
  Matches    4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db      99 AGYLTKGAGLNE 110
           |:||: :::|::
Qy       1 ASYLSTSSSLDD 12


RESULT    9
ID   US-08-459-174-2     STANDARD;     PRT;   213 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08459174
XX
CC    Sequence 2, Application US/08459174
CC    Patent No. 5710031
CC     GENERAL INFORMATION:
CC       APPLICANT:  Gaffney, Thomas D.
CC       APPLICANT:  Lam, Stephen T.
CC       APPLICANT:  Ligon, James M.
CC       APPLICANT:  Hill, Dwight S.
CC       APPLICANT:  Stein, Jeffrey I.
CC       APPLICANT:  Howell, Charles R.
CC       APPLICANT:  Becker, J. Ole
CC       TITLE OF INVENTION:  Gene Activating Element
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CIBA-GEIGY Corporation
CC         STREET:  7 Skyline Drive
CC         CITY:  Hawthorne
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10532
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/459,174
CC         FILING DATE:  02-JUN-1995
CC         CLASSIFICATION:  536
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/287,442
CC         FILING DATE:  08-AUG-1994
CC         APPLICATION NUMBER:  US 08/087,636
CC         FILING DATE:  01-JUL-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/908,284
CC         FILING DATE:  02-JUL-1992
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/570,184
CC         FILING DATE:  08-AUG-1990
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Elmer, James Scott
CC         REGISTRATION NUMBER:  36,129
CC         REFERENCE/DOCKET NUMBER:  CGC 1506/CIP4
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  919-541-8614
CC         TELEFAX:  919-541-8689
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  213 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   213 AA; 23364 MW; 236956 CN;

  Query Match           68.1%;  Score 47;  DB 1;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 7.38e+01;
  Matches     4;  Conservative     7;  Mismatches   1;  Indels   0;  Gaps   0;

Db      99 AGYLTKGAGLNE 110
           |:||: :::|::
Qy       1 ASYLSTSSSLDD 12


RESULT   10
ID   US-08-460-298-2     STANDARD;     PRT;   213 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08460298
XX
CC    Sequence 2, Application US/08460298
CC    Patent No. 5686283
CC     GENERAL INFORMATION:
CC       APPLICANT:  Gaffney, Thomas D.
CC       APPLICANT:  Lam, Stephen T.
CC       APPLICANT:  Ligon, James M.
CC       APPLICANT:  Hill, Dwight S.
CC       APPLICANT:  Stein, Jeffrey I.
CC       APPLICANT:  Howell, Charles R.
CC       APPLICANT:  Becker, J. Ole
CC       TITLE OF INVENTION:  Gene Activating Element
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CIBA-GEIGY Corporation
CC         STREET:  7 Skyline Drive
CC         CITY:  Hawthorne
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10532
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/460,298
CC         FILING DATE:  02-JUN-1995
CC         CLASSIFICATION:  536
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/287,442
CC         FILING DATE:  08-AUG-1994
CC         APPLICATION NUMBER:  US 08/087,636
CC         FILING DATE:  01-JUL-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/908,284
CC         FILING DATE:  02-JUL-1992
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/570,184
CC         FILING DATE:  08-AUG-1990
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Elmer, James Scott
CC         REGISTRATION NUMBER:  36,129
CC         REFERENCE/DOCKET NUMBER:  CGC 1506/CIP4
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  919-541-8614
CC         TELEFAX:  919-541-8689
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  213 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   213 AA; 23364 MW; 236956 CN;

  Query Match           68.1%;  Score 47;  DB 1;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 7.38e+01;
  Matches     4;  Conservative     7;  Mismatches   1;  Indels   0;  Gaps   0;

Db      99 AGYLTKGAGLNE 110
```

```
          |:||: :::|::
Qy      1 ASYLSTSSSLDD 12


RESULT   11
ID   US-08-761-258-3     STANDARD;     PRT;   213 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08761258
XX
CC    Sequence 3, Application US/08761258
CC    Patent No. 5756087
CC     GENERAL INFORMATION:
CC       APPLICANT:  Ligon, James M.
CC       APPLICANT:  Hill, Dwight S.
CC       APPLICANT:  Lam, Stephen T.
CC       APPLICANT:  Gaffney, Thomas D.
CC       APPLICANT:  Torkewitz, Nancy
CC       TITLE OF INVENTION:  Genetically Modified Pseudomonas Strains
CC       TITLE OF INVENTION:  with Enhanced Biocontrol Activity
CC       NUMBER OF SEQUENCES:  11
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Ciba-Geigy Corporation
CC         STREET:  520 White Plains Road, P.O. Box 2005
CC         CITY:  Tarrytown
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10591
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/761,258
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Meigs, J. Timothy
CC         REGISTRATION NUMBER:  38,241
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (919) 541-8587
CC         TELEFAX:  (919) 541-8689
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  213 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   213 AA; 23364 MW; 236956 CN;

  Query Match           68.1%;  Score 47;  DB 2;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 7.38e+01;
  Matches     4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db     99 AGYLTKGAGLNE 110
          |:||: :::|::
Qy      1 ASYLSTSSSLDD 12


RESULT   12
ID   PCT-US93-06300A-2     STANDARD;     PRT;   213 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9306300A
XX
CC    Sequence 2, Application PC/TUS9306300A
CC     GENERAL INFORMATION:
CC       APPLICANT:  CIBA-GEIGY AG
CC       APPLICANT:  Klybeckstrasse 141
CC       APPLICANT:  4002 Basle
CC       APPLICANT:  Switzerland
CC       APPLICANT:  125 Tradescant Road
CC       APPLICANT:  Chapel Hill, NC  27514
CC       APPLICANT:  USA
CC       APPLICANT:  8900 Jeanew Court
CC       APPLICANT:  Raleigh, NC  27613
CC       APPLICANT:  USA
CC       APPLICANT:  Hill, Dwight Steven
CC       APPLICANT:  311 Melanie Lane
CC       APPLICANT:  Cary, NC  27511
CC       APPLICANT:  USA
CC       APPLICANT:  Stein, Jeffrey I.
CC       APPLICANT:  3725 Surry Trail
CC       APPLICANT:  Hillsborough, NC  27278
CC       APPLICANT:  USA
CC       APPLICANT:  Howell, Charles R.
CC       APPLICANT:  805 Avondale
CC       APPLICANT:  Bryan, TX  77802
CC       APPLICANT:  USA
CC       APPLICANT:  Becker, J. Ole
CC       APPLICANT:  6164 Osevego
CC       APPLICANT:  Riverside, CA  92506
CC       APPLICANT:  USA
CC       APPLICANT:  Ligon, James M.
CC       APPLICANT:  120 Marquette Drive
CC       APPLICANT:  Cary, NC  27513
CC       APPLICANT:  USA
CC       TITLE OF INVENTION:  Gene Activating Element
CC       NUMBER OF SEQUENCES:  7
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CIBA-GEIGY Corporation
CC         STREET:  7 Skyline Drive
CC         CITY:  Hawthorne
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10532
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/06300A
CC         FILING DATE:  02-JUL-1993
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/908,284
CC         FILING DATE:  02-JUL-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Spruill, W. Murray
CC         REGISTRATION NUMBER:  32,943
CC         REFERENCE/DOCKET NUMBER:  S-18210/A/CGC1506/PC
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (919)541-8615
CC         TELEFAX:  (919)541-8689
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  213 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   213 AA; 23364 MW; 236956 CN;

  Query Match           68.1%;  Score 47;  DB 3;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 7.38e+01;
  Matches     4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db     99 AGYLTKGAGLNE 110
          |:||: :::|::
```

```
Qy      1 ASYLSTSSSLDD 12


RESULT   13
ID   US-08-468-540B-2     STANDARD;     PRT;   20 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08468540B
XX
CC    Sequence 2, Application US/08468540B
CC    Patent No. 5858980
CC     GENERAL INFORMATION:
CC       APPLICANT:  Weiner, Howard
CC       APPLICANT:  Hafler, David
CC       APPLICANT:  Miller, Ariel
CC       APPLICANT:  Al-Sabbagh, Ahmad
CC       TITLE OF INVENTION:  SUPPRESSION OF T-CELL PROLIFERATION
CC       TITLE OF INVENTION:  USING PEPTIDE FRAGMENTS OF MYELIN BASIC PROTEIN
CC       NUMBER OF SEQUENCES:  25
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Darby & Darby P.C.
CC         STREET:  805 Third Avenue
CC         CITY:  New York
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10022
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/468,540B
CC         FILING DATE:  06-JUN-1995
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Jacobs, Seth H
CC         REGISTRATION NUMBER:  32,140
CC         REFERENCE/DOCKET NUMBER:
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-527-7700
CC         TELEFAX:
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  20 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  No. 5858980e
SQ   SEQUENCE   20 AA; 2160 MW; 2047 CN;

  Query Match           66.7%;  Score 46;  DB 2;  Length 20;
  Best Local Similarity  45.5%;  Pred. No. 9.47e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db      2 SKYLATASTMD 12
          : ||:|:|::|
Qy      1 ASYLSTSSSLD 11


RESULT   14
ID   US-08-787-547-33     STANDARD;     PRT;   21 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 33, Application US/08787547
XX
CC    Sequence 33, Application US/08787547
CC    Patent No. 5783567
CC     GENERAL INFORMATION:
CC       APPLICANT:  Hedley, Mary Lynne
CC       APPLICANT:  Curley, Joanne M.
CC       APPLICANT:  Langer, Robert S.
CC       TITLE OF INVENTION:  MICROPARTICLES FOR DELIVERY
CC       TITLE OF INVENTION:  OF NUCLEIC ACID
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson, P.C.
CC         STREET:  225 Franklin Street
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  US
CC         ZIP:  02110-2804
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  Windows95
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/787,547
CC         FILING DATE:  22-JAN-1997
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Fraser, Janis K.
CC         REGISTRATION NUMBER:  34,819
CC         REFERENCE/DOCKET NUMBER:  08191/003001
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  617-542-5070
CC         TELEFAX:  617-542-8906
CC         TELEX:  200154
CC     INFORMATION FOR SEQ ID NO:  33:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  21 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   21 AA; 2312 MW; 2012 CN;

  Query Match           66.7%;  Score 46;  DB 2;  Length 21;
  Best Local Similarity  45.5%;  Pred. No. 9.47e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db      6 SKYLATASTMD 16
          : ||:|:|::|
Qy      1 ASYLSTSSSLD 11


RESULT   15
ID   US-08-327-357A-1     STANDARD;     PRT;   170 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08327357A
XX
CC    Sequence 1, Application US/08327357A
CC    Patent No. 5817629
CC     GENERAL INFORMATION:
CC       APPLICANT:  WARREN, Kenneth G.
CC       APPLICANT:  CATZ, Ingrid
CC       TITLE OF INVENTION:  PEPTIDE SPECIFICITY OF ANTI-MYELIN BASIC
CC       TITLE OF INVENTION:  PROTEIN AND THE ADMINISTRATION OF MYELIN BASIC PROTEIN
CC       TITLE OF INVENTION:  PEPTIDES TO MUTIPLE SCLEROSIS PATIENTS
```

```
CC       NUMBER OF SEQUENCES:  1
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Venable, Baetjer, Howard & Civiletti
CC         STREET:  1201 New York Avenue, N.W., Suite 1000
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20005
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/327,357A
CC         FILING DATE:  21-OCT-1994
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/798,099
CC         FILING DATE:  27-NOV-1991
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  CA 2,053,799-0
CC         FILING DATE:  22-OCT-1991
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Ihnen, Jeffrey L.
CC         REGISTRATION NUMBER:  28,957
CC         REFERENCE/DOCKET NUMBER:  27052-115469
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-962-4810
CC         TELEFAX:  202-962-8300
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  170 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Homo sapiens
CC       IMMEDIATE SOURCE:
CC         CLONE:  human myelin basic protein
SQ   SEQUENCE   170 AA; 18459 MW; 143992 CN;

  Query Match           66.7%;  Score 46;  DB 2;  Length 170;
  Best Local Similarity  45.5%;  Pred. No. 9.47e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     12 SKYLATASTMD 22
          : ||:|:|::|
        1 ASYLSTSSSLD 11


Search completed: Thu Sep  2 12:37:11 1999
Job time : 9 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:35:24 1999; MasPar time 3.47 Seconds
138.694 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-30
Description: (1-12) from US08599226.pep
Perfect Score: 69
Sequence: 1 ASYLSTSSSLDD 12

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.804; Variance 27.163; scale 0.876

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	55	79.7	377	2	F71520	hypothetical protein	3.98e-01
2	51	73.9	471	2	C71439	hypothetical protein	3.11e+00
3	50	72.5	293	2	D70108	conserved hypothetica	5.12e+00
4	49	71.0	345	1	VABPA7	major capsid protein	8.36e+00
5	49	71.0	381	2	G71906	probable transcriptio	8.36e+00
6	49	71.0	398	1	VBBPA7	minor capsid protein	8.36e+00
7	49	71.0	437	2	JC5459	inulin fructotransfer	8.36e+00
8	49	71.0	650	2	F70974	probable acrA1 protei	8.36e+00
9	48	69.6	239	2	S25204	srmX protein - Strept	1.35e+01
10	48	69.6	396	2	C49904	L-lactate dehydrogena	1.35e+01
11	48	69.6	483	2	A55033	keratin 12 - mouse	1.35e+01
12	48	69.6	687	2	A41905	ferric vibriobactin r	1.35e+01
13	48	69.6	750	2	S55180	phospholipase D homol	1.35e+01
14	47	68.1	213	2	A38222	regulatory protein ga	2.18e+01
15	47	68.1	333	2	S42424	hypothetical protein	2.18e+01
16	47	68.1	471	2	S08325	flavonol 3-O-glucosyl	2.18e+01
17	47	68.1	471	2	S01052	flavonol 3-O-glucosyl	2.18e+01
18	47	68.1	471	2	S01037	flavonol 3-O-glucosyl	2.18e+01
19	47	68.1	578	2	S26577	beta-hemolysin - Aero	2.18e+01
20	47	68.1	684	2	G70744	hypothetical protein	2.18e+01
21	47	68.1	792	2	C71250	probable cation-trans	2.18e+01
22	46	66.7	115	2	D32055	nifW protein - Azotob	3.47e+01
23	46	66.7	128	1	MBRTS	myelin basic protein	3.47e+01
24	46	66.7	132	2	H70759	hypothetical protein	3.47e+01
25	46	66.7	167	2	A37246	myelin basic protein	3.47e+01
26	46	66.7	171	1	MBCZB	myelin basic protein	3.47e+01
27	46	66.7	197	1	MBHUB	myelin basic protein	3.47e+01
28	46	66.7	236	2	S60941	probable membrane pro	3.47e+01
29	46	66.7	301	2	S13095	P34 protein - Rickett	3.47e+01
30	46	66.7	328	1	MBMSB	golli-myelin basic pr	3.47e+01
31	46	66.7	347	1	VABPA3	major capsid protein	3.47e+01
32	46	66.7	381	2	G64607	response regulator -	3.47e+01
33	46	66.7	411	2	S45318	keratin 12 - rabbit (	3.47e+01
34	46	66.7	433	1	VBBPA3	minor capsid protein	3.47e+01
35	46	66.7	433	2	S04665	minor capsid protein	3.47e+01
36	46	66.7	467	2	S10364	flagellar hook-associ	3.47e+01
37	46	66.7	564	2	S57124	CTP synthase (EC 6.3.	3.47e+01
38	46	66.7	832	2	S71788	P/CAF protein - human	3.47e+01
39	46	66.7	1091	2	A58532	glial cell membrane g	3.47e+01
40	46	66.7	3924	2	S37431	ankyrin 2, neuronal l	3.47e+01
41	45	65.2	147	2	A69198	hypothetical protein	5.49e+01
42	45	65.2	416	2	H64984	hypothetical 43.4 kD	5.49e+01
43	45	65.2	543	2	S56830	probable purine nucle	5.49e+01
44	45	65.2	561	2	A31994	keratin 10, type I, e	5.49e+01
45	45	65.2	1298	2	A64157	hypothetical protein	5.49e+01

ALIGNMENTS

RESULT 1
ENTRY F71520 #type complete
TITLE hypothetical protein yprS - Chlamydia trachomatis (serotype D, strain UW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
ACCESSIONS F71520
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession F71520
##status preliminary
##molecule_type DNA
##residues 1-377 ##label ARN
##cross-references GB:AE001312; GB:AE001273; NID:g3328812; PID:g3328818
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene yprS
SUMMARY #length 377 #molecular-weight 41449 #checksum 9699

Query Match 79.7%; Score 55; DB 2; Length 377;
Best Local Similarity 66.7%; Pred. No. 3.98e-01;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 324 ASYLSSSPSVED 335
|||||:|:|::|
Qy 1 ASYLSTSSSLDD 12

RESULT 2
ENTRY C71439 #type complete
TITLE hypothetical protein - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
#variety columbia
DATE 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
ACCESSIONS C71439
REFERENCE A71400
#authors Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.;

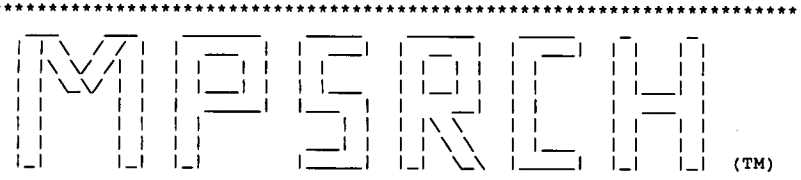
```
                  Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel,
                  K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.;
                  Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.;
                  Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.;
                  Lecharny, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.;
                  Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger,
                  M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
                  James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka,
                  A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.;
                  Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.;
                  Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.;
                  Rechman, S.; Ansorge, W.; Cooke, R.; Berger, C.; Delseny,
                  M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
                  Schueller, C.; Chalwatzis, N.
   #journal       Nature (1998) 391:485-488
   #title         Analysis of 1.9 Mb of contiguous sequence from chromosome 4
                    of Arabidopsis thaliana.
   #cross-references MUID:98121113
   #accession     C71439
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues      1-471 ##label BEV
      ##cross-references GB:Z97342; NID:g2245031; PID:e327038; PID:g2245065
GENETICS
   #map_position 4COP9-4G3845
SUMMARY           #length 471  #molecular-weight 52785  #checksum 7455

  Query Match             73.9%;  Score 51;  DB 2;  Length 471;
  Best Local Similarity  63.6%;  Pred. No. 3.11e+00;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     19 AGYLTSSSDLD 29
          |:||::|| ||
Qy      1 ASYLSTSSSLD 11


RESULT    3
ENTRY             D70108     #type complete
TITLE             conserved hypothetical protein BB0068 - Lyme disease
                    spirochete
ORGANISM          #formal_name Borrelia burgdorferi #common_name Lyme disease
                    spirochete
DATE              13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                    05-Jun-1998
ACCESSIONS        D70108
REFERENCE         A70100
   #authors       Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
                    Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
                    Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
                    J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
                    Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
                    Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
                    J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
                    Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
                    K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
   #journal       Nature (1997) 390:580-586
   #title         Genomic sequence of a Lyme disease spirochaete, Borrelia
                    burgdorferi.
   #cross-references MUID:98065943
   #accession     D70108
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues      1-293 ##label KLE
      ##cross-references GB:AE001120; GB:AE000783; NID:g2687951; PID:g2687956;
                          TIGR:BB0068
      ##experimental_source strain B31
SUMMARY           #length 293  #molecular-weight 33278  #checksum 5223

  Query Match             72.5%;  Score 50;  DB 2;  Length 293;
  Best Local Similarity  60.0%;  Pred. No. 5.12e+00;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    198 AYLSTPNSLE 207
          :||||::||:
Qy      2 SYLSTSSSLD 11


RESULT    4
ENTRY             VABPA7     #type complete
TITLE             major capsid protein 10A - phage T7
ORGANISM          #formal_name phage T7
   #note          host Escherichia coli
DATE              13-Jun-1983 #sequence_revision 30-Sep-1990 #text_change
                    26-Feb-1999
ACCESSIONS        A04344; S42325
REFERENCE         A94615
   #authors       Dunn, J.J.; Thompson, K.
   #submission    submitted to the Nucleic Acid Sequence Database, September
                    1982
   #accession     A04344
      ##molecule_type DNA
      ##residues      1-345 ##label DUN
REFERENCE         S42283
   #authors       Dunn, J.J.; Studier, F.W.
   #journal       J. Mol. Biol. (1983) 166:477-535
   #title         Complete nucleotide sequence of bacteriophage T7 DNA and the
                    locations of T7 genetic elements.
   #cross-references MUID:83241725
   #accession     S42325
      ##molecule_type DNA
      ##residues      1-345 ##label DUW
      ##cross-references EMBL:V01146; NID:g431187; PID:g15604
      ##note          the authors did not translate the codon for residue 1
GENETICS
   #gene          10A
   #map_position 57.51-60.49
CLASSIFICATION    #superfamily phage T7 major capsid protein 10A
KEYWORDS          capsid protein
SUMMARY           #length 345  #molecular-weight 36545  #checksum 8044

  Query Match             71.0%;  Score 49;  DB 1;  Length 345;
  Best Local Similarity  50.0%;  Pred. No. 8.36e+00;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     71 AAYLAPGENLDD 82
          |:||: : :|||
Qy      1 ASYLSTSSSLDD 12


RESULT    5
ENTRY             G71906     #type complete
TITLE             probable transcription regulator - Helicobacter pylori
                    (strain J99)
ORGANISM          #formal_name Helicobacter pylori
   #variety       strain J99
DATE              12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                    12-Feb-1999
ACCESSIONS        G71906
REFERENCE         A71800
   #authors       Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                    Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                    B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                    Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                    Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                    G.F.; Trust, T.J.
   #journal       Nature (1999) 397:176-180
   #title         Genomic sequence comparison of two unrelated isolates of the
                    human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession     G71906
      ##status         preliminary
      ##molecule_type DNA
      ##residues      1-381 ##label ARN
      ##cross-references GB:AE001496; GB:AE001439; NID:g4155191; PID:g4155196
```

```
      ##experimental_source strain J99
GENETICS
   #gene          jhp0643
SUMMARY           #length 381  #molecular-weight 43475  #checksum 2767

  Query Match           71.0%;  Score 49;  DB 2;  Length 381;
  Best Local Similarity  45.5%;  Pred. No. 8.36e+00;
  Matches     5;  Conservative     5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    135 SFLATSKALEE 145
          |:|:|| :|::
Qy      2 SYLSTSSSLDD 12


RESULT    6
ENTRY             VBBPA7      #type complete
TITLE             minor capsid protein 10B - phage T7
ORGANISM          #formal_name phage T7
   #note          host Escherichia coli
DATE              30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change
                    26-Feb-1999
ACCESSIONS        B04344; S42326
REFERENCE         A94615
   #authors       Dunn, J.J.; Thompson, K.
   #submission    submitted to the Nucleic Acid Sequence Database, September
                    1982
   #accession     B04344
      ##molecule_type DNA
      ##residues      1-398 ##label DUN
REFERENCE         S42283
   #authors       Dunn, J.J.; Studier, F.W.
   #journal       J. Mol. Biol. (1983) 166:477-535
   #title         Complete nucleotide sequence of bacteriophage T7 DNA and the
                    locations of T7 genetic elements.
   #cross-references MUID:83241725
   #accession     S42326
      ##status        preliminary; translation not shown
      ##molecule_type DNA
      ##residues      1-398 ##label DU2
      ##cross-references EMBL:V01146; NID:g431187; PID:g431193
GENETICS
   #gene          10B
   #map_position 57.51-60.49
   #note          translation of the nucleotide sequence involves a -1
                    frameshift within codon 341
CLASSIFICATION    #superfamily phage T7 major capsid protein 10A
KEYWORDS          capsid protein
SUMMARY           #length 398  #molecular-weight 41830  #checksum 5020

  Query Match           71.0%;  Score 49;  DB 1;  Length 398;
  Best Local Similarity  50.0%;  Pred. No. 8.36e+00;
  Matches     6;  Conservative     4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     71 AAYLAPGENLDD 82
          |:||: : :|||
Qy      1 ASYLSTSSSLDD 12


RESULT    7
ENTRY             JC5459      #type complete
TITLE             inulin fructotransferase (depolymerizing) (EC 2.4.1.93)
                    precursor - Arthrobacter sp.
ALTERNATE_NAMES   Inulase II; Inulin fructotransferase (DFA-III-producing)
ORGANISM          #formal_name Arthrobacter sp.
DATE              17-Jun-1997 #sequence_revision 18-Jul-1997 #text_change
                    13-Nov-1998
ACCESSIONS        JC5459
REFERENCE         JC5459
   #authors       Sakurai, H.; Yokota, A.; Tomita, F.
   #journal       Biosci. Biotechnol. Biochem. (1997) 61:87-92
   #title         Molecular cloning of an inulin fructotransferase
                    (depolymerizing) gene from Arthrobacter sp. H65-7 and its
                    expression in Escherichia coli.
   #cross-references MUID:97179800
   #accession     JC5459
      ##molecule_type DNA
      ##residues      1-437 ##label SAK
      ##cross-references DDBJ:D84399; NID:g1906791; PID:d1019710; PID:g1906792
      ##experimental_source strain H65-7
COMMENT           This enzyme is thermotolerant and metal ions resistant protein. It
                    is also stable for the industrial preparation of
                    di-D-fructofuranose 1,2':2,3' dianhydride. It converts inulin
                    into di-D-fructofuranose 1,2':2,3' dianhydride through
                    intramolecular transfructosylation.
GENETICS
   #gene          ift
KEYWORDS          glycosyltransferase; hexosyltransferase
FEATURE
   1-32                #domain signal sequence #status predicted #label SIG\
   33-437              #product inulin fructotransferase (depolymerizing)
                         #status predicted #label MAT
SUMMARY           #length 437  #molecular-weight 46475  #checksum 9304

  Query Match           71.0%;  Score 49;  DB 2;  Length 437;
  Best Local Similarity  50.0%;  Pred. No. 8.36e+00;
  Matches     6;  Conservative     6;  Mismatches   0;  Indels   0;  Gaps   0;

Db    331 PPFLGTSNGLDD 342
          :::|:||::|||
Qy      1 ASYLSTSSSLDD 12


RESULT    8
ENTRY             F70974      #type complete
TITLE             probable acrA1 protein - Mycobacterium tuberculosis (strain
                    H37RV)
ORGANISM          #formal_name Mycobacterium tuberculosis
DATE              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                    17-Jul-1998
ACCESSIONS        F70974
REFERENCE         A70500
   #authors       Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                    Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal       Nature (1998) 393:537-544
   #title         Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
   #cross-references MUID:98295987
   #accession     F70974
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-650 ##label COL
      ##cross-references GB:AL009198; GB:AL123456; NID:g3242262; PID:e1202309;
                          PID:g2661670
      ##experimental_source strain H37Rv
GENETICS
   #gene          acrA1
SUMMARY           #length 650  #molecular-weight 70939  #checksum 8551

  Query Match           71.0%;  Score 49;  DB 2;  Length 650;
  Best Local Similarity  54.5%;  Pred. No. 8.36e+00;
  Matches     6;  Conservative     4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    501 SSYLPTKAALD 511
          :|||:| ::||
Qy      1 ASYLSTSSSLD 11
```

```
RESULT    9
ENTRY             S25204     #type complete
TITLE             srmX protein - Streptomyces ambofaciens
ORGANISM          #formal_name Streptomyces ambofaciens
DATE              28-May-1993 #sequence_revision 28-May-1993 #text_change
                    01-May-1998
ACCESSIONS        S25204; S21599
REFERENCE         S25202
   #authors       Geistlich, M.; Losick, R.; Turner, J.R.; Rao, R.N.
   #journal       Mol. Microbiol. (1992) 6:2019-2029
   #title         Characterization of a novel regulatory gene governing the
                    expression of a polyketide synthase gene in Streptomyces
                    ambofaciens.
   #accession     S25204
      ##molecule_type DNA
      ##residues       1-239 ##label GEI
      ##cross-references EMBL:X63451; NID:g46699; PID:g46702
GENETICS
   #gene          srmX
CLASSIFICATION    #superfamily bioC homology
FEATURE
   39-139              #domain bioC homology #label BIOC
SUMMARY           #length 239  #molecular-weight 26493  #checksum 3795

  Query Match           69.6%;  Score 48;  DB 2;  Length 239;
  Best Local Similarity  45.5%;  Pred. No. 1.35e+01;
  Matches     5;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    110 GYLETTADLED 120
          :|| |:: |:|
Qy      2 SYLSTSSSLDD 12


RESULT   10
ENTRY             C49904     #type complete
TITLE             L-lactate dehydrogenase (EC 1.1.1.27), FMN-dependent -
                    Escherichia coli
ALTERNATE_NAMES   lctD protein
ORGANISM          #formal_name Escherichia coli
DATE              11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change
                    17-Mar-1999
ACCESSIONS        C49904; G65160; S47826
REFERENCE         A49904
   #authors       Dong, J.M.; Taylor, J.S.; Latour, D.J.; Iuchi, S.; Lin,
                    E.C.C.
   #journal       J. Bacteriol. (1993) 175:6671-6678
   #title         Three overlapping lct genes involved in L-lactate utilization
                    by Escherichia coli.
   #cross-references MUID:94012541
   #accession     C49904
      ##status        preliminary
      ##molecule_type DNA
      ##residues       1-396 ##label DON
      ##cross-references GB:L13970; NID:g404692; PID:g404695
REFERENCE         A64720
   #authors       Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                    Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                    Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                    Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                    Y.
   #journal       Science (1997) 277:1453-1462
   #title         The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession     G65160
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-396 ##label BLAT
      ##cross-references GB:AE000438; GB:U00096; NID:g2367251; PID:g1790033;
                        UWGP:b3605
      ##experimental_source strain K-12, substrain MG1655
REFERENCE         S47666
   #authors       Plunkett, G.
   #submission    submitted to the EMBL Data Library, March 1994
   #accession     S47826
      ##status        preliminary
      ##molecule_type DNA
      ##residues       1-118,'X',120-396 ##label PLU
      ##cross-references EMBL:U00039; NID:g466582; PID:g466743
GENETICS
   #gene          lctD
CLASSIFICATION    #superfamily (S)-2-hydroxy-acid oxidase; (S)-2-hydroxy-acid
                    oxidase homology
KEYWORDS          FMN; oxidoreductase
FEATURE
   1-321               #domain (S)-2-hydroxy-acid oxidase homology #label 2HY\
   275                 #active_site His #status predicted
SUMMARY           #length 396  #molecular-weight 42728  #checksum 9705

  Query Match           69.6%;  Score 48;  DB 2;  Length 396;
  Best Local Similarity  33.3%;  Pred. No. 1.35e+01;
  Matches     4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db    208 SAYLGKPTGLED 219
          ::||: :::|:|
Qy      1 ASYLSTSSSLDD 12


RESULT   11
ENTRY             A55033     #type complete
TITLE             keratin 12 - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change
                    17-Mar-1999
ACCESSIONS        A55033
REFERENCE         A55033
   #authors       Liu, C.Y.; Zhu, G.; Converse, R.; Kao, C.W.C.; Nakamura, H.;
                    Tseng, S.C.G.; Mui, M.M.; Seyer, J.; Justice, M.J.; Stech,
                    M.E.; Hansen, G.M.; Kao, W.W.Y.
   #journal       J. Biol. Chem. (1994) 269:24627-24636
   #title         Characterization and chromosomal localization of the
                    cornea-specific murine keratin gene Krt1.12.
   #cross-references MUID:95014223
   #accession     A55033
      ##status        preliminary
      ##molecule_type DNA
      ##residues       1-483 ##label LIU
      ##cross-references GB:U08095; NID:g565659; PID:g565660
      ##note          authors translated the codon ATC for residue 225 as Thr,
                       and GCG for residue 388 as Arg
CLASSIFICATION    #superfamily cytoskeletal keratin
SUMMARY           #length 483  #molecular-weight 52010  #checksum 7

  Query Match           69.6%;  Score 48;  DB 2;  Length 483;
  Best Local Similarity  50.0%;  Pred. No. 1.35e+01;
  Matches     6;  Conservative    3;  Mismatches   3;  Indels   0;  Gaps   0;

Db    131 ASYLGKVRSLEE 142
          ||||:   ||::
Qy      1 ASYLSTSSSLDD 12


RESULT   12
ENTRY             A41905     #type complete
TITLE             ferric vibriobactin receptor ViuA - Vibrio cholerae
ALTERNATE_NAMES   iron-siderophore complex receptor ViuA
ORGANISM          #formal_name Vibrio cholerae
DATE              04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
                    20-Mar-1998
ACCESSIONS        A41905
REFERENCE         A41905
   #authors       Butterton, J.R.; Stoebner, J.A.; Payne, S.M.; Calderwood,
                    S.B.
   #journal       J. Bacteriol. (1992) 174:3729-3738
   #title         Cloning, sequencing, and transcriptional regulation of viuA,
                    the gene encoding the ferric vibriobactin receptor of
                    Vibrio cholerae.
```

```
   #cross-references MUID:92276356
   #contents     0395
   #accession    A41905
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-687 ##label BUT
      ##cross-references GB:U11759; GB:M90461; NID:g517206; PID:g531822
      ##note          sequence extracted from NCBI backbone (NCBIN:104700,
                       NCBIP:104701)
SUMMARY          #length 687  #molecular-weight 76412  #checksum 2802

  Query Match           69.6%;  Score 48;  DB 2;  Length 687;
  Best Local Similarity  60.0%;  Pred. No. 1.35e+01;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    662 YLSTNNTLDQ 671
          ||||:::||:
Qy      3 YLSTSSSLDD 12


RESULT   13
ENTRY            S55180     #type complete
TITLE            phospholipase D homolog YJL132w - yeast (Saccharomyces
                   cerevisiae)
ALTERNATE_NAMES  protein J0678
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
                   05-Jun-1998
ACCESSIONS       S55180; S56914; S71665
REFERENCE        S55159
   #authors      Katsoulou, C.; Tzermia, M.; Alexandraki, D.
   #submission   submitted to the EMBL Data Library, May 1995
   #description  The complete sequence of a 40.7 kb segment located on the
                   left arm of yeast chromosome X identified 13 known genes
                   and revealed 13 new open reading frames including
                   homologues to other yeast hypothetical proteins.
   #accession    S55180
      ##molecule_type DNA
      ##residues      1-750 ##label KAT
      ##cross-references EMBL:X87371; NID:g854542; PID:g854564
REFERENCE        S56912
   #authors      Katsoulou, C.; Tzermia, M.; Alexandraki, D.
   #submission   submitted to the Protein Sequence Database, September 1995
   #accession    S56914
      ##molecule_type DNA
      ##residues      1-750 ##label KAW
      ##cross-references EMBL:Z49407; NID:g1008335; PID:g1008336; MIPS:YJL132w
REFERENCE        S71643
   #authors      Katsoulou, C.; Tzermia, M.; Tavernarakis, N.; Alexandraki, D.
   #journal      Yeast (1996) 12:787-797
   #title        Sequence analysis of a 40.7 kb segment from the left arm of
                   yeast chromosome X reveals 14 known genes and 13 new open
                   reading frames including homologues of genes clustered on
                   the right arm of chromosome XI.
   #accession    S71665
      ##status        nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues      1-750 ##label KAF
      ##cross-references EMBL:X87371; NID:g854542; PID:g854564
GENETICS
   #map_position 10L
SUMMARY          #length 750  #molecular-weight 84466  #checksum 4191

  Query Match           69.6%;  Score 48;  DB 2;  Length 750;
  Best Local Similarity  50.0%;  Pred. No. 1.35e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    354 YLAVSAPLED 363
          ||: |::|:|
Qy      3 YLSTSSSLDD 12


RESULT   14
ENTRY            A38222     #type complete
TITLE            regulatory protein gacA - Pseudomonas fluorescens
ORGANISM         #formal_name Pseudomonas fluorescens
DATE             31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
                   05-Dec-1997
ACCESSIONS       A38222
REFERENCE        A38222
   #authors      Laville, J.; Voisard, C.; Keel, C.; Maurhofer, M.; Defago,
                   G.; Haas, D.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1562-1566
   #title        Global control in Pseudomonas fluorescens mediating
                   antibiotic synthesis and suppression of black root rot of
                   tobacco.
   #cross-references MUID:92179223
   #contents     CHA0
   #accession    A38222
      ##molecule_type DNA
      ##residues      1-213 ##label LAV
      ##note          this sequence is inconsistent with the nucleotide
                       translation
      ##note          sequence extracted from NCBI backbone (NCBIN:85369,
                       NCBIP:85373)
GENETICS
   #gene         gacA
CLASSIFICATION   #superfamily regulatory protein comA; response regulator
                   homology
KEYWORDS         phosphoprotein
FEATURE
   4-115             #domain response regulator homology #label RRH\
   54                #binding_site phosphate (Asp) (covalent) #status
                       predicted
SUMMARY          #length 213  #molecular-weight 23663  #checksum 7318

  Query Match           68.1%;  Score 47;  DB 2;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 2.18e+01;
  Matches     4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db     99 AGYLTKGAGLNE 110
          |:||: :::|::
Qy      1 ASYLSTSSSLDD 12


RESULT   15
ENTRY            S42424     #type complete
TITLE            hypothetical protein y - Mycobacterium smegmatis
ORGANISM         #formal_name Mycobacterium smegmatis
DATE             07-May-1998 #sequence_revision 15-May-1998 #text_change
                   15-May-1998
ACCESSIONS       S42424; S31804
REFERENCE        S42421
   #authors      Cirillo, J.D.; Weisbrod, T.R.; Pascopella, L.; Bloom, B.R.;
                   Jacobs Jr., W.R.
   #journal      Mol. Microbiol. (1994) 11:629-639
   #title        Isolation and characterization of the aspartokinase and
                   aspartate semialdehyde dehydrogenase operon from
                   mycobacteria.
   #accession    S42424
      ##molecule_type DNA
      ##residues      1-333 ##label CIR
      ##cross-references EMBL:Z17372; NID:g44506; PID:g581353
      ##note          the authors translated the initiation codon GTG for
                       residue 1 as Val
GENETICS
   #start_codon  GTG
SUMMARY          #length 333  #molecular-weight 35881  #checksum 4453

  Query Match           68.1%;  Score 47;  DB 2;  Length 333;
  Best Local Similarity  41.7%;  Pred. No. 2.18e+01;
  Matches     5;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db    290 GGYISPASTLDE 301
          ::|:| :|:||:
Qy      1 ASYLSTSSSLDD 12
```

Search completed: Thu Sep 2 12:35:41 1999
Job time : 17 secs.



MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:36:23 1999; MasPar time 4.43 Seconds
147.813 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-30
Description: (1-12) from US08599226.pep
Perfect Score: 69
Sequence: 1 ASYLSTSSSLDD 12

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 23.424; Variance 25.983; scale 0.902

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	55	79.7	377	2	O84397	HYPOTHETICAL 41.4 KD P	3.17e-01
2	53	76.8	127	14	O11696	NUCLEOPROTEIN (FRAGMEN	9.36e-01
3	51	73.9	471	10	O23552	HYPOTHETICAL 52.8 KD P	2.69e+00
4	50	72.5	293	2	O51095	CONSERVED HYPOTHETICAL	4.51e+00
5	49	71.0	188	10	O81650	THYMIDYLATE KINASE (FR	7.51e+00
6	49	71.0	384	2	O11197	HYPOTHETICAL 41.9 KD P	7.51e+00
7	49	71.0	437	2	O08305	INULIN FRUCTOTRANSFERA	7.51e+00
8	49	71.0	650	2	O50417	MULTI-FUNCTIONAL ENZYM	7.51e+00
9	49	71.0	1022	6	Q28628	A-KINASE ANCHORING PRO	7.51e+00
10	48	69.6	160	5	Q27302	GLOBIN.	1.24e+01
11	48	69.6	239	2	Q00510	SRMX PROTEIN.	1.24e+01
12	48	69.6	484	5	Q23173	W05E10.1 PROTEIN.	1.24e+01
13	48	69.6	649	3	O60167	PROTEIN COMPLEX ASSEMB	1.24e+01
14	48	69.6	1758	5	Q22830	SIMILAR TO HUMAN SREBP	1.24e+01
15	47	68.1	112	2	O68946	DINITROGENASE 3 DELTA	2.03e+01
16	47	68.1	213	2	O69157	RESPONSE REGULATOR.	2.03e+01
17	47	68.1	471	5	O61986	DYSTROPHIN-LIKE PROTEI	2.03e+01
18	47	68.1	696	5	Q23264	ZC373.1 PROTEIN (FRAGM	2.03e+01
19	47	68.1	792	2	O83999	CATION-TRANSPORTING AT	2.03e+01
20	47	68.1	931	10	O04026	HYPOTHETICAL 104.6 KD	2.03e+01
21	47	68.1	1548	5	O01583	SIMILAR TO SERINE/THRE	2.03e+01
22	46	66.7	149	11	Q61836	MYELIN BASIC PROTEIN (	3.31e+01
23	46	66.7	160	5	Q27430	GLOBIN.	3.31e+01
24	46	66.7	195	11	Q01585	MYELIN BASIC PROTEIN (	3.31e+01
25	46	66.7	197	4	Q15339	GOLLI-MBP.	3.31e+01
26	46	66.7	236	3	Q12282	CHROMOSOME XV READING	3.31e+01
27	46	66.7	250	5	Q23509	COSMID ZK470.	3.31e+01
28	46	66.7	250	11	Q03139	MYELIN BASIC PROTEIN (	3.31e+01
29	46	66.7	258	5	Q19850	COSMID F28B12.	3.31e+01
30	46	66.7	304	4	Q15340	GOLLI-MBP.	3.31e+01
31	46	66.7	381	2	O25408	RESPONSE REGULATOR.	3.31e+01
32	46	66.7	831	5	Q21574	M28.8 PROTEIN.	3.31e+01
33	46	66.7	832	4	Q92831	P/CAF.	3.31e+01
34	46	66.7	857	10	Q39392	SRK29 PROTEIN KINASE.	3.31e+01
35	46	66.7	1091	11	P70193	MEMBRANE GLYCOPROTEIN.	3.31e+01
36	46	66.7	1128	1	O51999	REPI PROTEIN.	3.31e+01
37	46	66.7	1128	1	O52009	REPJ PROTEIN.	3.31e+01
38	46	66.7	2697	5	O01438	SIMILARITY TO THE P13/	3.31e+01
39	45	65.2	132	2	O68943	DINITROGENASE 3 DELTA	5.33e+01
40	45	65.2	423	13	O93256	KERATIN-19.	5.33e+01
41	45	65.2	473	4	Q16402	TYPE I KERATIN 16.	5.33e+01
42	45	65.2	583	5	Q18517	C39B10.1 PROTEIN.	5.33e+01
43	45	65.2	852	10	O23243	BETA-GALACTOSIDASE.	5.33e+01
44	45	65.2	853	10	Q42150	BETA-GALACTOSIDASE LIK	5.33e+01
45	45	65.2	1043	5	O01757	SIMILAR TO ACHLYA AMBI	5.33e+01

ALIGNMENTS

RESULT 1
ID O84397 PRELIMINARY; PRT; 377 AA.
AC O84397;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 41.4 KD PROTEIN.
GN YPRS.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis.";
RL SCIENCE 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AE001312; G3328818; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 377 AA; 41449 MW; 2B79B006 CRC32;

Query Match 79.7%; Score 55; DB 2; Length 377;
Best Local Similarity 66.7%; Pred. No. 3.17e-01;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Db    324 ASYLSSSPSVED 335
          |||||:|:|::|
Qy      1 ASYLSTSSSLDD 12
```

RESULT 2
ID O11696 PRELIMINARY; PRT; 127 AA.
AC O11696;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)

```
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NUCLEOPROTEIN (FRAGMENT).
OS   MEASLES VIRUS (SUBACUTE SCLEROSE PANENCEPHALITIS VIRUS).
OC   VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC   PARAMYXOVIRIDAE; PARAMYXOVIRINAE; MORBILLIVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=92-E;
RX   MEDLINE; 97278133.
RA   YAMAGUCHI S.;
RT   "Identification of three lineages of wild measles virus by nucleotide
RT   sequence analysis of N, P, M, F, and L genes in Japan.";
RL   J. MED. VIROL. 52:113-120(1997).
DR   EMBL; D87487; D1020995; -.
KW   NUCLEOPROTEIN.
FT   NON_TER       1      1
SQ   SEQUENCE   127 AA;  13950 MW;  42D75A2C CRC32;

  Query Match           76.8%;  Score 53;  DB 14;  Length 127;
  Best Local Similarity  63.6%;  Pred. No. 9.36e-01;
  Matches    7;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     66 AAYLPTSTPLD 76
          |:||:||::||
Qy      1 ASYLSTSSSLD 11


RESULT    3
ID   O23552      PRELIMINARY;      PRT;   471 AA.
AC   O23552;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 52.8 KD PROTEIN.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BEVAN M., STIEKEMA W., MURPHY G., WAMBUTT R., POHL T., TERRYN N.,
RA   KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
RA   PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,
RA   JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
RA   SCHUELLER C., CHALWATZIS N.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z97342; E327038; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   471 AA;  52785 MW;  4EBA8315 CRC32;

  Query Match           73.9%;  Score 51;  DB 10;  Length 471;
  Best Local Similarity  63.6%;  Pred. No. 2.69e+00;
  Matches    7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     19 AGYLTSSSDLD 29
          |:||::|| ||
Qy      1 ASYLSTSSSLD 11


RESULT    4
ID   O51095      PRELIMINARY;      PRT;   293 AA.
AC   O51095;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   BB0068.
OS   BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC   BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE; 98065943.
RA   FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA   LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA   DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA   PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA   VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA   UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA   GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA   SMITH H.O., VENTER J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   NATURE 390:580-586(1997).
DR   EMBL; AE001120; G2687956; -.
DR   TIGR; BB0068; -.
SQ   SEQUENCE   293 AA;  33278 MW;  3FBBB9E2 CRC32;

  Query Match           72.5%;  Score 50;  DB 2;  Length 293;
  Best Local Similarity  60.0%;  Pred. No. 4.51e+00;
  Matches    6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    198 AYLSTPNSLE 207
          :||||::||:
Qy      2 SYLSTSSSLD 11


RESULT    5
ID   O81650      PRELIMINARY;      PRT;   188 AA.
AC   O81650;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   THYMIDYLATE KINASE (FRAGMENT).
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ULLAH H., ROBERTSON N., FITES R.C.;
RT   "Plant thymidylate kinase mRNA.";
RL   SUBMITTED (AUG-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF081570; G3493131; -.
FT   NON_TER       1      1
SQ   SEQUENCE   188 AA;  21248 MW;  C306E7B8 CRC32;

  Query Match           71.0%;  Score 49;  DB 10;  Length 188;
  Best Local Similarity  58.3%;  Pred. No. 7.51e+00;
  Matches    7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db     43 SAYLSNKSQLDD 54
          ::|||  | |||
Qy      1 ASYLSTSSSLDD 12


RESULT    6
ID   Q11197      PRELIMINARY;      PRT;   384 AA.
AC   Q11197;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 41.9 KD PROTEIN IN CMA1 3'REGION (ORF2).
OS   MYCOBACTERIUM TUBERCULOSIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RA;
RX   MEDLINE; 95327699.
RA   YUAN Y., LEE R.E., BESRA G.S., BELISLE J.T., BARRY C.E. III;
RT   "Identification of a gene involved in the biosynthesis of
```

```
RT   cyclopropanated mycolic acids in Mycobacterium tuberculosis.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 92:6630-6634(1995).
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       FAMILY (SDR).
DR   EMBL; U27357; G886103; -.
DR   PFAM; PF00106; adh_short; 1.
KW   HYPOTHETICAL PROTEIN; OXIDOREDUCTASE.
SQ   SEQUENCE   384 AA;  41995 MW;  9FF8465A CRC32;

  Query Match           71.0%;  Score 49;  DB 2;  Length 384;
  Best Local Similarity  54.5%;  Pred. No. 7.51e+00;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    235 SSYLPTKAALD 245
          :|||:| ::||
Qy      1 ASYLSTSSSLD 11


RESULT    7
     O08305      PRELIMINARY;      PRT;   437 AA.
     O08305;
     01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   INULIN FRUCTOTRANSFERASE (EC 2.4.1.93).
GN   IFT.
OS   ARTHROBACTER SP.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; MICROCOCCINEAE; MICROCOCCACEAE; ARTHROBACTER.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97179800.
RA   SAKURAI H., YOKOTA A., TOMITA F.;
RT   "Molecular cloning of an Inulin fructotransferase (depolymerizing)
RT   gene from Arthrobacter sp. H65-7 and its expression in Escherichia
RT   coli.";
RL   BIOSCI. BIOTECHNOL. BIOCHEM. 61:87-92(1997).
DR   EMBL; D84399; D1019710; -.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   437 AA;  46475 MW;  DC74B68C CRC32;

  Query Match           71.0%;  Score 49;  DB 2;  Length 437;
  Best Local Similarity  50.0%;  Pred. No. 7.51e+00;
  Matches     6;  Conservative    6;  Mismatches   0;  Indels   0;  Gaps   0;

Db    331 PPFLGTSNGLDD 342
          :::|:||::|||
Qy      1 ASYLSTSSSLDD 12


RESULT    8
ID   O50417      PRELIMINARY;      PRT;   650 AA.
AC   O50417;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MULTI-FUNCTIONAL ENZYME.
GN   MTV004.49.
OS   MYCOBACTERIUM TUBERCULOSIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   OLIVER K., SKELTON J., BADCOCK K., CHURCHER C.M., HARRIS D.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR   EMBL; AL009198; E1202309; -.
SQ   SEQUENCE   650 AA;  70939 MW;  1E552EA5 CRC32;

  Query Match           71.0%;  Score 49;  DB 2;  Length 650;
  Best Local Similarity  54.5%;  Pred. No. 7.51e+00;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    501 SSYLPTKAALD 511
          :|||:| ::||
Qy      1 ASYLSTSSSLD 11


RESULT    9
ID   Q28628      PRELIMINARY;      PRT;  1022 AA.
AC   Q28628;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   A-KINASE ANCHORING PROTEIN AKAP120.
OS   ORYCTOLAGUS CUNICULUS (RABBIT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=STOMACH;
RX   MEDLINE; 97220389.
RA   DRANSFIELD D.T., YEH J.L., BRADFORD A.J., GOLDENRING J.R.;
RT   "Identification and characterization of a novel A-kinase-anchoring
RT   protein (AKAP120) from rabbit gastric parietal cells.";
RL   BIOCHEM. J. 322:0-0(0).
DR   EMBL; U26360; G3599584; -.
SQ   SEQUENCE  1022 AA;  116903 MW;  98BD43C2 CRC32;

  Query Match           71.0%;  Score 49;  DB 6;  Length 1022;
  Best Local Similarity  72.7%;  Pred. No. 7.51e+00;
  Matches     8;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    935 AYLSTISSLKD 945
          :|||| ||| |
Qy      2 SYLSTSSSLDD 12


RESULT   10
ID   Q27302      PRELIMINARY;      PRT;   160 AA.
AC   Q27302;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   GLOBIN.
GN   CBG1.
OS   CAENORHABDITIS BRIGGSAE.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KLOEK A.P., GOLDBERG D.E.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U48291; G1203913; -.
DR   EMBL; U48290; G1203913; JOINED.
DR   EMBL; U48289; G1203909; -.
DR   PFAM; PF00042; globin; 1.
SQ   SEQUENCE   160 AA;  18591 MW;  618F4F46 CRC32;

  Query Match           69.6%;  Score 48;  DB 5;  Length 160;
```

```
  Best Local Similarity  41.7%;  Pred. No. 1.24e+01;
  Matches     5;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db    119 TGYLESTGSLND 130
          ::|| :::||:|
Qy      1 ASYLSTSSSLDD 12


RESULT   11
ID   Q00510      PRELIMINARY;      PRT;   239 AA.
AC   Q00510;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SRMX PROTEIN.
GN   SRMX.
OS   STREPTOMYCES AMBOFACIENS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; STREPTOMYCINEAE; STREPTOMYCETACEAE; STREPTOMYCES.
     [1]
     SEQUENCE FROM N.A.
RC   STRAIN=BES2281;
RX   MEDLINE; 92374852.
RA   GEISTLICH M., LOSICK R., TURNER J.R., RAU R.N.;
RT   "Characterization of a novel regulatory gene governing the expression
RT   of a polyketide synthase gene in Streptomyces ambofaciens.";
RL   MOL. MICROBIOL. 6:2019-2029(1992).
DR   EMBL; X63451; G46702; -.
SQ   SEQUENCE   239 AA;  26493 MW;  45A8C51A CRC32;

  Query Match           69.6%;  Score 48;  DB 2;  Length 239;
  Best Local Similarity  45.5%;  Pred. No. 1.24e+01;
  Matches     5;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    110 GYLETTADLED 120
          :|| |:: |:|
Qy      2 SYLSTSSSLDD 12


RESULT   12
ID   Q23173      PRELIMINARY;      PRT;   484 AA.
AC   Q23173;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   W05E10.1 PROTEIN.
GN   W05E10.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MORTIMORE B.;
RL   SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z77670; E1350053; -.
SQ   SEQUENCE   484 AA;  53896 MW;  EC00CF7B CRC32;

  Query Match           69.6%;  Score 48;  DB 5;  Length 484;
  Best Local Similarity  45.5%;  Pred. No. 1.24e+01;
  Matches     5;  Conservative    6;  Mismatches   0;  Indels   0;  Gaps   0;

Db    150 AYVAASSTLED 160
          :|:::||:|:|
Qy      2 SYLSTSSSLDD 12


RESULT   13
ID   O60167      PRELIMINARY;      PRT;   649 AA.
AC   O60167;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   PROTEIN COMPLEX ASSEMBLY PROTEIN.
GN   SPBC19F8.03C.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   BECK A., REINHARDT R., WOOD V., RAJANDREAM M.A., BARRELL B.G.;
RL   SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AL023594; E1293401; -.
SQ   SEQUENCE   649 AA;  72985 MW;  9C207DB2 CRC32;

  Query Match           69.6%;  Score 48;  DB 3;  Length 649;
  Best Local Similarity  60.0%;  Pred. No. 1.24e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    234 NYLSTARSLE 243
          :|||| ||:
Qy      2 SYLSTSSSLD 11


RESULT   14
ID   Q22830      PRELIMINARY;      PRT;  1758 AA.
AC   Q22830;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO HUMAN SREBP-2 BASIC-HELIX-LOOP-HELIX-LEUCINE ZIPPER
DE   TRANSCRIPTION FACTOR.
GN   T27B1.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   WU X., LE T.;
RL   SUBMITTED (DEC-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
```

```
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U41020; G1086678; -.
SQ   SEQUENCE   1758 AA;  198406 MW;  A58AC1E4 CRC32;

  Query Match           69.6%;  Score 48;  DB 5;  Length 1758;
  Best Local Similarity  41.7%;  Pred. No. 1.24e+01;
  Matches     5;  Conservative    4;  Mismatches   3;  Indels   0;  Gaps   0;

Db    962 SSYLAKAEDLDE 973
          :|||: :  ||:
Qy      1 ASYLSTSSSLDD 12


RESULT   15
ID   O68946      PRELIMINARY;      PRT;   112 AA.
AC   O68946;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
     DINITROGENASE 3 DELTA SUBUNIT.
     ANFG.
     AZOMONAS MACROCYTOGENES.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; AZOTOBACTERACEAE;
OC   AZOMONAS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LOVELESS T.M., BISHOP P.E.;
RL   SUBMITTED (APR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF058780; G3065908; -.
SQ   SEQUENCE   112 AA;  13273 MW;  5ABCB39C CRC32;

  Query Match           68.1%;  Score 47;  DB 2;  Length 112;
  Best Local Similarity  50.0%;  Pred. No. 2.03e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db     96 YLTISGSLNE 105
          ||: |:||::
Qy      3 YLSTSSSLDD 12


Search completed: Thu Sep  2 12:36:46 1999
Job time : 23 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:35:58 1999; MasPar time 2.17 Seconds
156.010 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-30
Description: (1-12) from US08599226.pep
Perfect Score: 69
Sequence: 1 ASYLSTSSSLDD 12

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 24.293; Variance 24.652; scale 0.985

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	49	71.0	345	1	VCAA_BPT7	MAJOR CAPSID PROTEIN 1	3.05e+00
2	49	71.0	352	1	YDH3_SCHPO	HYPOTHETICAL 39.7 KD P	3.05e+00
3	49	71.0	398	1	VCAB_BPT7	MINOR CAPSID PROTEIN 1	3.05e+00
4	48	69.6	396	1	LLDD_ECOLI	L-LACTATE DEHYDROGENAS	5.18e+00
5	48	69.6	483	1	K1CL_MOUSE	KERATIN, TYPE I CYTOSK	5.18e+00
6	48	69.6	687	1	VIUA_VIBCH	VIBRIOBACTIN RECEPTOR	5.18e+00
7	48	69.6	750	1	YJN2_YEAST	HYPOTHETICAL 84.5 KD P	5.18e+00
8	47	68.1	213	1	GACA_PSEFL	RESPONSE REGULATOR GAC	8.71e+00
9	47	68.1	333	1	YASD_MYCSM	HYPOTHETICAL 35.9 KD P	8.71e+00
10	47	68.1	471	1	UFO1_MAIZE	FLAVONOL 3-O-GLUCOSYLT	8.71e+00
11	47	68.1	471	1	UFO3_MAIZE	FLAVONOL 3-O-GLUCOSYLT	8.71e+00
12	47	68.1	471	1	UFO2_MAIZE	FLAVONOL 3-O-GLUCOSYLT	8.71e+00
13	47	68.1	578	1	HLY4_AERSA	HEMOLYSIN 4 PRECURSOR.	8.71e+00
14	47	68.1	684	1	YV18_MYCTU	HYPOTHETICAL 57.3 KD P	8.71e+00
15	46	66.7	115	1	NIFW_AZOVI	NIFW PROTEIN.	1.45e+01
16	46	66.7	127	1	MBP_RAT	MYELIN BASIC PROTEIN S	1.45e+01
17	46	66.7	132	1	YW07_MYCTU	HYPOTHETICAL 14.7 KD P	1.45e+01
18	46	66.7	167	1	MBP_CAVPO	MYELIN BASIC PROTEIN (	1.45e+01
19	46	66.7	171	1	MBP_PANTR	MYELIN BASIC PROTEIN (	1.45e+01
20	46	66.7	194	1	MBP_MOUSE	MYELIN BASIC PROTEIN (	1.45e+01
21	46	66.7	196	1	MBP_HUMAN	MYELIN BASIC PROTEIN (	1.45e+01
22	46	66.7	301	1	P34_RICRI	PROTEIN P34.	1.45e+01
23	46	66.7	347	1	VCAA_BPT3	MAJOR CAPSID PROTEIN 1	1.45e+01
24	46	66.7	411	1	K1CL_RABIT	KERATIN, TYPE I CYTOSK	1.45e+01
25	46	66.7	433	1	VCAB_BPT3	MINOR CAPSID PROTEIN 1	1.45e+01
26	46	66.7	466	1	FLID_SALTY	FLAGELLAR HOOK-ASSOCIA	1.45e+01
27	46	66.7	577	1	PYRH_YEAST	CTP SYNTHASE 2 (EC 6.3	1.45e+01
28	46	66.7	1839	1	ANKC_HUMAN	ANKYRIN, BRAIN VARIANT	1.45e+01
29	46	66.7	3924	1	ANKB_HUMAN	ANKYRIN, BRAIN VARIANT	1.45e+01
30	45	65.2	132	1	ANFG_AZOVI	NITROGENASE IRON-IRON	2.40e+01
31	45	65.2	204	1	UREG_STRSL	UREASE ACCESSORY PROTE	2.40e+01
32	45	65.2	401	1	K1CS_POTTR	KERATIN, TYPE I CYTOSK	2.40e+01
33	45	65.2	416	1	YEIJ_ECOLI	HYPOTHETICAL 43.4 KD P	2.40e+01
34	45	65.2	431	1	K1CQ_HUMAN	KERATIN, TYPE I CYTOSK	2.40e+01
35	45	65.2	456	1	K1CO_HUMAN	KERATIN, TYPE I CYTOSK	2.40e+01
36	45	65.2	469	1	K1CX_HUMAN	KERATIN, TYPE I CYTOSK	2.40e+01
37	45	65.2	471	1	K1CN_HUMAN	KERATIN, TYPE I CYTOSK	2.40e+01
38	45	65.2	494	1	K1CL_HUMAN	KERATIN, TYPE I CYTOSK	2.40e+01
39	45	65.2	526	1	K1CJ_BOVIN	KERATIN, TYPE I CYTOSK	2.40e+01
40	45	65.2	543	1	YJF8_YEAST	HYPOTHETICAL 60.8 KD P	2.40e+01
41	45	65.2	547	1	MERA_STAAU	MERCURIC REDUCTASE (EC	2.40e+01
42	45	65.2	593	1	K1CJ_HUMAN	KERATIN, TYPE I CYTOSK	2.40e+01
43	45	65.2	633	1	YIJ2_YEAST	HYPOTHETICAL 71.0 KD P	2.40e+01
44	45	65.2	753	1	YBHJ_ECOLI	HYPOTHETICAL 81.5 KD P	2.40e+01
45	45	65.2	1298	1	YTFN_HAEIN	HYPOTHETICAL PROTEIN H	2.40e+01

ALIGNMENTS

RESULT 1
ID VCAA_BPT7 STANDARD; PRT; 345 AA.
AC P19726; P03717;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE MAJOR CAPSID PROTEIN 10A.
GN 10.
OS BACTERIOPHAGE T7.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; PODOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83241725.
RA DUNN J.J., STUDIER F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT locations of T7 genetic elements.";
RL J. MOL. BIOL. 166:477-535(1983).
CC -!- SIMILARITY: TO THE T3 MAJOR CAPSID PROTEIN 10A.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; V01146; G15604; -.
DR PIR; A04344; VABPA7.
DR PIR; S42325; S42325.
KW COAT PROTEIN.
SQ SEQUENCE 345 AA; 36545 MW; D12C1BAE CRC32;

Query Match 71.0%; Score 49; DB 1; Length 345;
Best Local Similarity 50.0%; Pred. No. 3.05e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```
Db      71 AAYLAPGENLDD 82
           |:||: : :|||
Qy       1 ASYLSTSSSLDD 12
```

RESULT 2
ID YDH3_SCHPO STANDARD; PRT; 352 AA.
AC Q92348;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

```
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 39.7 KD PROTEIN C6G9.03C IN CHROMOSOME I.
GN   SPAC6G9.03C.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., CONNOR R.E.;
RL   SUBMITTED (OCT-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: SOME, TO YEAST YNL206C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z81317; E276610; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   352 AA;  39679 MW;  D92A9357 CRC32;

  Query Match           71.0%;  Score 49;  DB 1;  Length 352;
  Best Local Similarity  54.5%;  Pred. No. 3.05e+00;
  Matches    6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    303 SYLNEASSLEN 313
          |||: :|||::
Qy      2 SYLSTSSSLDD 12


RESULT    3
ID   VCAB_BPT7      STANDARD;      PRT;   398 AA.
AC   P19727; P03717;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE   MINOR CAPSID PROTEIN 10B.
GN   10.
OS   BACTERIOPHAGE T7.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; PODOVIRIDAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83241725.
RA   DUNN J.J., THOMPSON K.;
RT   "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT   locations of T7 genetic elements.";
RL   J. MOL. BIOL. 166:477-535(1983).
CC   -!- THE MINOR CAPSID PROTEIN 10B IS DUE TO A TRANSLATIONAL SHIFT TO
CC       THE -1 FRAME.
CC   -!- SIMILARITY: TO THE T3 MINOR CAPSID PROTEIN 10B, EXCEPT IN THE
CC       C-TERMINAL EXTENSIONS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01146; G431193; -.
DR   PIR; B04344; VBBPA7.
DR   PIR; S42326; S42326.
KW   COAT PROTEIN.
SQ   SEQUENCE   398 AA;  41830 MW;  88D410FF CRC32;

  Query Match           71.0%;  Score 49;  DB 1;  Length 398;
  Best Local Similarity  50.0%;  Pred. No. 3.05e+00;
  Matches    6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     71 AAYLAPGENLDD 82
          |:||: : :|||
Qy      1 ASYLSTSSSLDD 12


RESULT    4
ID   LLDD_ECOLI     STANDARD;      PRT;   396 AA.
AC   P33232;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   L-LACTATE DEHYDROGENASE (CYTOCHROME) (EC 1.1.2.3).
GN   LLDD OR LCTD.
OS   ESCHERICHIA COLI.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ESCHERICHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 94012541.
RA   DONG J.M., TAYLOR J.S., LATOUR D.J., IUCHI S., LIN E.C.C.;
RT   "Three overlapping lct genes involved in L-lactate utilization by
RT   Escherichia coli.";
RL   J. BACTERIOL. 175:6671-6678(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 94316500.
RA   SOFIA H.J., BURLAND V., DANIELS D.L., PLUNKETT G. III, BLATTNER F.R.;
RT   "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT   region from 76.0 to 81.5 minutes.";
RL   NUCLEIC ACIDS RES. 22:2576-2586(1994).
CC   -!- CATALYTIC ACTIVITY: (S)-LACTATE + 2 FERRICYTOCHROME C = PYRUVATE +
CC       2 FERROCYTOCHROME C.
CC   -!- COFACTOR: FMN.
CC   -!- INDUCTION: AEROBICALLY BY L-LACTATE.
CC   -!- SIMILARITY: BELONGS TO THE FMN-DEPENDENT ALPHA-HYDROXY ACID
CC       DEHYDROGENASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13970; G404695; -.
DR   EMBL; U00039; G466743; -.
DR   EMBL; AE000438; G1790033; -.
DR   PIR; C49904; C49904.
DR   ECOGENE; EG11963; LLDD.
DR   PROSITE; PS00557; FMN_HYDROXY_ACID_DH; 1.
DR   PFAM; PF01070; FMN_dh; 1.
DR   HSSP; P05414; 1GYL.
KW   OXIDOREDUCTASE; FLAVOPROTEIN; FMN.
FT   ACT_SITE     24     24       SUBSTRATE BINDING (BY SIMILARITY).
FT   ACT_SITE    129    129       SUBSTRATE BINDING (BY SIMILARITY).
FT   ACT_SITE    275    275       REMOVES THE SUBSTRATE ALPHA-PROTON AS THE
FT                                FIRST STEP IN CATALYSIS (BY SIMILARITY).
FT   ACT_SITE    278    278       SUBSTRATE BINDING (BY SIMILARITY).
SQ   SEQUENCE   396 AA;  42728 MW;  10BAD2D1 CRC32;

  Query Match           69.6%;  Score 48;  DB 1;  Length 396;
  Best Local Similarity  33.3%;  Pred. No. 5.18e+00;
  Matches    4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db    208 SAYLGKPTGLED 219
          ::||: :::|:|
Qy      1 ASYLSTSSSLDD 12
```

```
RESULT    5
ID   K1CL_MOUSE     STANDARD;      PRT;   483 AA.
AC   Q64291;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   KERATIN, TYPE I CYTOSKELETAL 12 (CYTOKERATIN 12).
GN   KRT12 OR KRT1-12 OR KRT1.12.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=CORNEA;
RX   MEDLINE; 94139368.
RA   LIU C.-Y., ZHU G., WESTERHAUSEN-LARSON A., CONVERSE R., KAO C.W.-C.,
RA   SUN T.-T., KAO W.W.-Y.;
RT   "Cornea-specific expression of K12 keratin during mouse development.";
RL   CURR. EYE RES. 12:963-974(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ; TISSUE=CORNEA;
RX   MEDLINE; 95014223.
RA   LIU C.-Y., ZHU G., CONVERSE R., KAO C.W.-C., NAKAMURA H.,
RA   TSENG S.C.-G., MUI M.-M., SEYER J., JUSTICE M.J., STECH M.E.,
RA   HANSEN G.M., KAO W.W.-Y.;
RT   "Characterization and chromosomal localization of the cornea-specific
RT   murine keratin gene Krt1.12.";
RL   J. BIOL. CHEM. 269:24627-24636(1994).
CC   -!- FUNCTION: MAY PLAY A UNIQUE ROLE IN MAINTAINING THE NORMAL CORNEAL
CC       EPITHELIAL FUNCTION.
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC       KERATIN 3 ASSOCIATES WITH KERATIN 12.
CC   -!- TISSUE SPECIFICITY: CORNEA-SPECIFIC. ASSOCIATED MAINLY WITH ALL
CC       LAYERS OF THE CENTRAL CORNEAL EPITHELIUM AND ALSO FOUND IN THE
CC       SUPRABASAL LIMBAL EPITHELIUM.
CC   -!- THERE ARE TWO TYPES OF CYTOSKELETAL AND MICROFIBRILLAR KERATIN,
CC       I (ACIDIC) AND II (NEUTRAL TO BASIC) (40-55 AND 56-70 KILODALTONS,
CC       RESPECTIVELY).
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENT FORMS MAY BE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U02880; G414540; -.
DR   EMBL; U08095; G565660; -.
DR   MGD; MGI:96687; KRT1-12.
DR   PROSITE; PS00226; IF; 1.
DR   PFAM; PF00038; filament; 1.
KW   INTERMEDIATE FILAMENT; COILED COIL; HEPTAD REPEAT PATTERN; KERATIN;
KW   ALTERNATIVE SPLICING.
FT   DOMAIN        1    118       HEAD.
FT   DOMAIN      119    428       ROD.
FT   DOMAIN      429    483       TAIL.
FT   DOMAIN      119    154       COIL 1A.
FT   DOMAIN      158    175       LINKER 1.
FT   DOMAIN      176    267       COIL 1B.
FT   DOMAIN      268    290       LINKER 12.
FT   DOMAIN      291    428       COIL 2.
FT   VARSPLIC    211    212       YE -> KL (IN K12-ALT9).
FT   VARSPLIC    213    483       MISSING (IN K12-ALT9).
SQ   SEQUENCE   483 AA;  52010 MW;  54DB7635 CRC32;

  Query Match           69.6%;  Score 48;  DB 1;  Length 483;
  Best Local Similarity 50.0%;  Pred. No. 5.18e+00;
  Matches     6;  Conservative    3;  Mismatches   3;  Indels   0;  Gaps   0;

Db    131 ASYLGKVRSLEE 142
          ||||:   ||::
Qy      1 ASYLSTSSSLDD 12


RESULT    6
ID   VIUA_VIBCH     STANDARD;      PRT;   687 AA.
AC   Q00964;
DT   01-APR-1993 (REL. 25, CREATED)
DT   01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   VIBRIOBACTIN RECEPTOR PRECURSOR.
GN   VIUA.
OS   VIBRIO CHOLERAE.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OGAWA 395;
RX   MEDLINE; 92276356.
RA   BUTTERTON J.R., STOEBNER J.A., PAYNE S.M., CALDERWOOD S.B.;
RT   "Cloning, sequencing, and transcriptional regulation of viuA, the
RT   gene encoding the ferric vibriobactin receptor of Vibrio cholerae.";
RL   J. BACTERIOL. 174:3729-3738(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OGAWA 395;
RA   LIAO W.J., CHOI M.H., BUTTERTON J.R.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: RECEPTOR FOR FERRIC VIBRIOBACTIN.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF030977; G2641147; -.
DR   PIR; A41905; A41905.
DR   PFAM; PF00593; TonB_boxC; 1.
KW   SIGNAL; RECEPTOR; OUTER MEMBRANE; IRON TRANSPORT.
FT   SIGNAL        1     37
FT   CHAIN        38    687       VIBRIOBACTIN RECEPTOR.
SQ   SEQUENCE   687 AA;  76413 MW;  525D3D49 CRC32;

  Query Match           69.6%;  Score 48;  DB 1;  Length 687;
  Best Local Similarity 60.0%;  Pred. No. 5.18e+00;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    662 YLSTNNTLDQ 671
          ||||:::||:
Qy      3 YLSTSSSLDD 12


RESULT    7
ID   YJN2_YEAST     STANDARD;      PRT;   750 AA.
AC   P47014;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 84.5 KD PROTEIN IN MRS3-URA2 INTERGENIC REGION.
GN   YJL132W OR J0678.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE; 96408771.
```

```
RA   KATSOULOU C., TZERMIA M., TAVERNARAKIS N., ALEXANDRAKI D.;
RT   "Sequence analysis of a 40.7 kb segment from the left arm of yeast
RT   chromosome X reveals 14 known genes and 13 new open reading frames
RT   including homologues of genes clustered on the right arm of
RT   chromosome XI.";
RL   YEAST 12:787-797(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: SOME, TO MAMMALIAN PHOSPHATIDYLINOSITOL-GLYCAN-
CC       SPECIFIC PHOSPHOLIPASE D (EC 3.1.4.50).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X87371; G854564; -.
DR   EMBL; Z49407; G1008336; -.
KW   HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT   TRANSMEM      1     21       POTENTIAL.
FT   TRANSMEM    465    485       POTENTIAL.
FT   TRANSMEM    586    606       POTENTIAL.
SQ   SEQUENCE   750 AA;  84466 MW;  5DD6D701 CRC32;

  Query Match           69.6%;  Score 48;  DB 1;  Length 750;
  Best Local Similarity  50.0%;  Pred. No. 5.18e+00;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    354 YLAVSAPLED 363
          ||: |::|:|
Qy      3 YLSTSSSLDD 12


RESULT    8
ID   GACA_PSEFL     STANDARD;      PRT;   213 AA.
AC   P32967;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   RESPONSE REGULATOR GACA (GLOBAL ACTIVATOR) (GLOBAL ANTIBIOTIC AND
DE   CYANIDE CONTROL PROTEIN).
GN   GACA.
OS   PSEUDOMONAS FLUORESCENS.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC   PSEUDOMONAS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CHA0;
RX   MEDLINE; 92179223.
RA   LAVILLE J., VOISARD C.P., KEEL C., MAURHOFER M., DIFAGO G.,
RA   HAAS D.;
RT   "Global control in Pseudomonas fluorescens mediating antibiotic
RT   synthesis and suppression of black root rot of tobacco.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 89:1562-1566(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL915;
RX   MEDLINE; 94355677.
RA   GAFFNEY T.D., LAM S.T., LIGON J., GATES K., FRAZELLE A., MAIO J.,
RA   HILL S., GOODWIN S., TORKEWITZ N., ALLSHOUSE A.M., KEMPF H.J.,
RA   BECKER J.O.;
RT   "Global regulation of expression of antifungal factors by a
RT   Pseudomonas fluorescens biological control strain.";
RL   MOL. PLANT MICROBE INTERACT. 7:455-463(1994).
CC   -!- FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM.
CC       INVOLVED IN THE SYNTHESIS OF THE ANTIFUNGAL FACTORS CYANIDE AND
CC       2,4-DIACETYLPHLOROGLUCINOL.
CC   -!- SIMILARITY: TO OTHER BACTERIAL REGULATORY PROTEINS INVOLVED IN
CC       SIGNAL TRANSDUCTION.
CC   -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80913; G551930; -.
DR   EMBL; L29642; G472403; -.
DR   PROSITE; PS00622; HTH_LUXR_FAMILY; 1.
DR   PFAM; PF00072; response_reg; 1.
DR   PFAM; PF00196; GerE; 1.
DR   HSSP; P10957; 1RNL.
KW   TRANSCRIPTION REGULATION; DNA-BINDING; SENSORY TRANSDUCTION;
KW   PHOSPHORYLATION.
FT   DOMAIN        1     99       RECEIVER DOMAIN.
FT   MOD_RES      54     54       PHOSPHORYLATION (BY SIMILARITY).
FT   DNA_BIND    166    185       H-T-H MOTIF (POTENTIAL).
FT   VARIANT      90     90       P -> L (IN THE SPONTANEOUS PLEIOTROPIC
FT                                MUTANT BL915-1).
FT   VARIANT     182    182       T -> I (IN THE SPONTANEOUS PLEIOTROPIC
FT                                MUTANT BL915-2).
FT   CONFLICT     49     49       Y -> D (IN REF. 2).
FT   CONFLICT    213    213       L -> A (IN REF. 2).
SQ   SEQUENCE   213 AA;  23454 MW;  42DC9AC2 CRC32;

  Query Match           68.1%;  Score 47;  DB 1;  Length 213;
  Best Local Similarity  33.3%;  Pred. No. 8.71e+00;
  Matches     4;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db     99 AGYLTKGAGLNE 110
          |:||: :::|::
Qy      1 ASYLSTSSSLDD 12


RESULT    9
ID   YASD_MYCSM     STANDARD;      PRT;   333 AA.
AC   P41402;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 35.9 KD PROTEIN IN ASD 3'REGION (ORFY).
OS   MYCOBACTERIUM SMEGMATIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 607 / MC(2)6;
RX   MEDLINE; 94254720.
RA   CIRILLO J.D., WEISBROD T.R., PASCOPELLA L., BLOOM B.R.,
RA   JACOBS W.R. JR.;
RT   "Isolation and characterization of the aspartokinase and aspartate
RT   semialdehyde dehydrogenase operon from mycobacteria.";
RL   MOL. MICROBIOL. 11:629-639(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z17372; G581353; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   333 AA;  35881 MW;  52FC5B23 CRC32;

  Query Match           68.1%;  Score 47;  DB 1;  Length 333;
  Best Local Similarity  41.7%;  Pred. No. 8.71e+00;
  Matches     5;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db    290 GGYISPASTLDE 301
          ::|:| :|:||:
Qy      1 ASYLSTSSSLDD 12


RESULT   10
ID   UFO1_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16166;
DT   01-APR-1990 (REL. 14, CREATED) .
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MCC ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88284304.
RA   RALSTON E.J., ENGLISH J.J., DOONER H.K.;
RT   "Sequence of three bronze alleles of maize and correlation with the
RT   genetic fine structure.";
RL   GENETICS 119:185-197(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13500; G1030071; -.
DR   EMBL; X07940; G22205; -.
DR   PIR; S01052; S01052.
DR   PIR; S08324; S08324.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48769 MW;  8AE03FD2 CRC32;

  Query Match           68.1%;  Score 47;  DB 1;  Length 471;
  Best Local Similarity  63.6%;  Pred. No. 8.71e+00;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     47 SFLSTASSLAQ 57
          |:|||:||| :
Qy      2 SYLSTSSSLDD 12


RESULT   11
ID   UFO3_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16167;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-W22 ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88284304.
RA   RALSTON E.J., ENGLISH J.J., DOONER H.K.;
RT   "Sequence of three bronze alleles of maize and correlation with the
RT   genetic fine structure.";
RL   GENETICS 119:185-197(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13502; G22506; -.
DR   EMBL; X07937; G22210; -.
DR   PIR; S01037; S01037.
DR   PIR; S08326; S08326.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48673 MW;  4A3C6193 CRC32;

  Query Match           68.1%;  Score 47;  DB 1;  Length 471;
  Best Local Similarity  63.6%;  Pred. No. 8.71e+00;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     47 SFLSTASSLAQ 57
          |:|||:||| :
Qy      2 SYLSTSSSLDD 12


RESULT   12
ID   UFO2_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16165;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MC2 ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
```

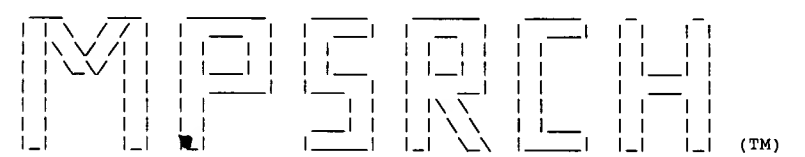
```
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13501; G295854; -.
DR   PIR; S08325; S08325.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48621 MW;  3158C5E0 CRC32;

  Query Match           68.1%;  Score 47;  DB 1;  Length 471;
  Best Local Similarity  63.6%;  Pred. No. 8.71e+00;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     47 SFLSTASSLAQ 57
          |:|||:||| :
Qy      2 SYLSTSSSLDD 12


RESULT   13
ID   HLY4_AERSA     STANDARD;      PRT;   578 AA.
AC   Q08677;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HEMOLYSIN 4 PRECURSOR.
GN   ASH4.
OS   AEROMONAS SALMONICIDA.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; AEROMONAS GROUP;
OC   AEROMONAS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=17-2;
RX   MEDLINE; 94142497.
RA   HIRONO I., AOKI T.;
RT   "Cloning and characterization of three hemolysin genes from Aeromonas
RT   salmonicida.";
RL   MICROB. PATHOG. 15:269-282(1993).
CC   -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC       CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC       DEFINED.
CC   -!- SIMILARITY: BELONGS TO THE AHH1/ASH4/HLYA/VVHA FAMILY OF
CC       HEMOLYSINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65049; G39012; -.
DR   PFAM; PF00652; Ricin_B_lectin; 1.
KW   HEMOLYSIS; TOXIN; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    578       HEMOLYSIN 4.
SQ   SEQUENCE   578 AA;  63400 MW;  8A23C8BC CRC32;

  Query Match           68.1%;  Score 47;  DB 1;  Length 578;
  Best Local Similarity  60.0%;  Pred. No. 8.71e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db     13 YLNASSWLDE 22
          ||::|| ||:
Qy      3 YLSTSSSLDD 12


RESULT   14
ID   YV18_MYCTU     STANDARD;      PRT;   684 AA.
AC   Q11157;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 57.3 KD PROTEIN GMC-TYPE OXIDOREDUCTASE CY20G9.18C.
GN   MTCY20G9.18C.
OS   MYCOBACTERIUM TUBERCULOSIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- COFACTOR: FAD FLAVOPROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z77162; E255032; -.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
DR   PROSITE; PS00624; GMC_OXRED_2; 1.
KW   HYPOTHETICAL PROTEIN; OXIDOREDUCTASE; FLAVOPROTEIN; FAD.
SQ   SEQUENCE   684 AA;  72253 MW;  F4A66D1F CRC32;

  Query Match           68.1%;  Score 47;  DB 1;  Length 684;
  Best Local Similarity  70.0%;  Pred. No. 8.71e+00;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    117 ASYLTTGRSL 126
          ||||:|: ||
Qy      1 ASYLSTSSSL 10


RESULT   15
ID   NIFW_AZOVI     STANDARD;      PRT;   115 AA.
AC   P14888;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NIFW PROTEIN.
GN   NIFW.
OS   AZOTOBACTER VINELANDII.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; AZOTOBACTERACEAE;
OC   AZOTOBACTER.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89123097.
RA   JACOBSON M.R., BRIGLE K.E., BENNETT L.T., SETTERQUIST R.A.,
RA   WILSON M.S., CASH V.L., BEYNON J., NEWTON W.E., DEAN D.R.;
RT   "Physical and genetic map of the major nif gene cluster from
RT   Azotobacter vinelandii.";
RL   J. BACTERIOL. 171:1017-1027(1989).
CC   -!- SIMILARITY: BELONGS TO THE NIFW FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M20568; G142365; -.
DR PIR; D32055; D32055.
KW NITROGEN FIXATION.
SQ SEQUENCE 115 AA; 13418 MW; 74E9995C CRC32;

Query Match 66.7%; Score 46; DB 1; Length 115;
Best Local Similarity 50.0%; Pred. No. 1.45e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 54 YLSKAGDLDE 63
||| :: ||:
Qy 3 YLSTSSSLDD 12

Search completed: Thu Sep 2 12:36:06 1999
Job time : 8 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:38:14 1999; MasPar time 3.62 Seconds
70.454 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-31
Description: (1-12) from US08599226.pep
Perfect Score: 82
Sequence: 1 ASYLSTSFSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.911; Variance 64.247; scale 0.279

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	82	100.0	12	27	W27590	Anti-TNF-alpha antibo	2.71e-01
2	67	81.7	12	27	W27588	Anti-TNF-alpha antibo	8.27e+00
3	64	78.0	12	27	W27593	Anti-TNF-alpha antibo	1.61e+01
4	61	74.4	12	27	W27591	Anti-TNF-alpha antibo	3.10e+01
5	61	74.4	12	27	W27592	Anti-TNF-alpha antibo	3.10e+01
6	61	74.4	121	27	W27569	Anti-TNF-alpha antibo	3.10e+01
7	55	67.1	12	27	W27586	Anti-TNF-alpha antibo	1.12e+02
8	55	67.1	12	27	W27589	Anti-TNF-alpha antibo	1.12e+02
9	55	67.1	12	27	W27587	Anti-TNF-alpha antibo	1.12e+02
10	55	67.1	141	9	R47205	Human/murine IL-1 chi	1.12e+02
11	55	67.1	147	9	R47208	Human/murine IL-1 chi	1.12e+02
12	55	67.1	147	9	R47211	Chimeric human/mouse	1.12e+02
13	54	65.9	277	30	W47089	Rat interleukin-1 bet	1.39e+02
14	53	64.6	783	35	W64560	Candida albicans olig	1.71e+02
15	52	63.4	306	15	R83040	Capsular polysacchari	2.11e+02
16	52	63.4	450	35	W71507	Helicobacter polypept	2.11e+02
17	51	62.2	180	23	W10875	FasL/LT-alpha hybrid.	2.59e+02
18	51	62.2	186	33	W55016	Amino acid sequence o	2.59e+02
19	51	62.2	278	33	W55014	Protein sequence of p	2.59e+02
20	51	62.2	303	39	W70843	Human Zcytor5 variant	2.59e+02
21	51	62.2	303	39	W70845	Human Zcytor5 variant	2.59e+02
22	51	62.2	350	33	W55015	Amino acid sequence o	2.59e+02
23	51	62.2	381	1	R05557	Transglutaminase type	2.59e+02
24	51	62.2	385	39	W70842	Human Zcytor5 variant	2.59e+02
25	51	62.2	385	39	W70841	Human Zcytor5 variant	2.59e+02
26	51	62.2	388	39	W70839	Human Zcytor5 variant	2.59e+02
27	51	62.2	389	39	W70844	Human Zcytor5 variant	2.59e+02
28	51	62.2	389	39	W70848	Human Zcytor5 variant	2.59e+02
29	51	62.2	389	39	W70849	Human Zcytor5 variant	2.59e+02
30	51	62.2	389	39	W70851	Human Zcytor5 variant	2.59e+02
31	51	62.2	389	39	W70850	Human Zcytor5 variant	2.59e+02
32	51	62.2	389	39	W70853	Human Zcytor5 variant	2.59e+02
33	51	62.2	389	39	W70852	Human Zcytor5 variant	2.59e+02
34	51	62.2	389	39	W70846	Human Zcytor5 variant	2.59e+02
35	51	62.2	389	39	W70847	Human Zcytor5 variant	2.59e+02
36	51	62.2	392	39	W70840	Human Zcytor5 variant	2.59e+02
37	51	62.2	408	34	W59805	Amino acid sequence o	2.59e+02
38	51	62.2	413	33	W55011	Novel haemopoietin re	2.59e+02
39	51	62.2	422	39	W70860	Human Zcytor5 protein	2.59e+02
40	51	62.2	425	39	W70861	Allelic varaint of hu	2.59e+02
41	51	62.2	425	39	W70862	Rat Zcytor5 protein.	2.59e+02
42	51	62.2	425	33	W55012	Novel haemopoietin re	2.59e+02
43	51	62.2	425	34	W59804	Nucleotide sequence o	2.59e+02
44	51	62.2	817	7	R34942	Human epidermal trans	2.59e+02
45	51	62.2	3396	8	R43662	DEN1-S275/90 (ECACC V	2.59e+02

ALIGNMENTS

RESULT 1
ID W27590 standard; peptide; 12 AA.
AC W27590;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          100.0%;  Score 82;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 2.71e-01;
  Matches    12;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsfsldy 12
          ||||||||||||
Qy      1 ASYLSTSFSLDY 12


RESULT    2
ID   W27588 standard; peptide; 12 AA.
AC   W27588;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          81.7%;  Score 67;  DB 27;  Length 12;
  Best Local Similarity 91.7%;  Pred. No. 8.27e+00;
  Matches    11;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 asylstsssldy 12
          ||||||| ||||
Qy      1 ASYLSTSFSLDY 12


RESULT    3
ID   W27593 standard; peptide; 12 AA.
AC   W27593;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 75; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          78.0%;  Score 64;  DB 27;  Length 12;
  Best Local Similarity 75.0%;  Pred. No. 1.61e+01;
  Matches     9;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 asylstassley 12
          ||||||: ||:|
Qy      1 ASYLSTSFSLDY 12


RESULT    4
ID   W27591 standard; peptide; 12 AA.
AC   W27591;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
```

KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 74.4%; Score 61; DB 27; Length 12;
Best Local Similarity 83.3%; Pred. No. 3.10e+01;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 asylstssslhy 12
||||||| || |
Qy 1 ASYLSTSFSLDY 12

RESULT 5
ID W27592 standard; peptide; 12 AA.
AC W27592;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 74.4%; Score 61; DB 27; Length 12;
Best Local Similarity 75.0%; Pred. No. 3.10e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1 asflstsssley 12
||:|||| ||:|
Qy 1 ASYLSTSFSLDY 12

RESULT 6
ID W27569 standard; Protein; 121 AA.
AC W27569;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain variable region.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody;
KW heavy chain; variable region; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
DR N-PSDB; T88404.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 16; Page 76; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain variable region.

CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 121 AA;

Query Match 74.4%; Score 61; DB 27; Length 121;
Best Local Similarity 81.8%; Pred. No. 3.10e+01;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 100 sylstassldy 110
|||||: ||||
Qy 2 SYLSTSFSLDY 12

RESULT 7
ID W27586 standard; peptide; 12 AA.
AC W27586;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 72; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 67.1%; Score 55; DB 27; Length 12;
Best Local Similarity 90.9%; Pred. No. 1.12e+02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 asylstsssld 11
||||||| |||
Qy 1 ASYLSTSFSLD 11

RESULT 8
ID W27589 standard; peptide; 12 AA.
AC W27589;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 67.1%; Score 55; DB 27; Length 12;
Best Local Similarity 90.9%; Pred. No. 1.12e+02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 asylstsssld 11
||||||| |||
Qy 1 ASYLSTSFSLD 11

```
RESULT    9
ID   W27587 standard; peptide; 12 AA.
AC   W27587;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           67.1%;  Score 55;  DB 27;  Length 12;
  Best Local Similarity  90.9%;  Pred. No. 1.12e+02;
  Matches    10;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 asylstsssld 11
          ||||||| |||
Qy      1 ASYLSTSFSLD 11


RESULT   10
ID   R47205 standard; Protein; 141 AA.
AC   R47205;
DT   09-AUG-1994  (first entry)
DE   Human/murine IL-1 chimeric antibody CL.
KW   Probe; chimeric; recombinant; antibody; human; interleukin-1; IL-1;
KW   light; L; chain; constant; region; variable; mouse; anti-human;
KW   graft; CDR; complementarity determining region; heavy; H;
KW   inflammatory disease; arteriosclerosis; detection;
KW   diffused intravascular coagulation; leukemia.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1..19
FT                   /note= "Signal peptide"
FT   protein         20..141
FT                   /note= "Mature protein"
PN   WO9402627-A.
PD   03-FEB-1994.
PF   08-JUL-1993; J00941.
PR   16-JUL-1992; JP-189248.
PA   (SAKA ) OTSUKA PHARM CO LTD.
PI   Hirai Y,  Nishida T,  Omoto Y,  Owens RJ;
DR   WPI; 94-048885/06.
DR   N-PSDB; Q56066.
PT   Mouse/human chimeric antibody against human interleukin-1 - for
PT   treatment of diseases in which production of interleukin-1 is
PT   abnormal, and for diagnostic imaging of interleukin-1 production
PT   sites in vivo
PS   Claim 1; Page 30-31; 58pp; Japanese.
CC   The sequences given in R47205-08 represent the light and heavy chain,
CC   variable and constant regions of a chimeric recombinant antibody
CC   against human interleukin-1 (IL-1).  The antibody has a light (L)
CC   chain in which the constant region is that of a human antibody and
CC   the variable region is from a mouse anti-human IL-1 antibody or is a
CC   mouse-human graft containing the CDR regions of mouse anti-human IL-1
CC   antibody, and a heavy (H) chain in which the constant region is that
CC   of a human antibody and the variable region is from a mouse anti-human
CC   IL-1 antibody or is a mouse-human graft containing the CDR regions of
CC   mouse anti-human IL-1 antibody,.  The chimeric antibody is used to
CC   treat diseases in which abnormal amounts of IL-1 are produced, eg.
CC   inflammatory disease, arteriosclerosis, diffused intravascular
CC   coagulation or leukemia.  It can also be labelled and used for
CC   diagnostic imaging of IL-1 producing sites in vivo.
SQ   Sequence   141 AA;

  Query Match           67.1%;  Score 55;  DB 9;  Length 141;
  Best Local Similarity  50.0%;  Pred. No. 1.12e+02;
  Matches     5;  Conservative    5;  Mismatches   0;  Indels   0;  Gaps   0;

Db    121 yfgssyaldy 130
          |:::|::|||
Qy      3 YLSTSFSLDY 12


RESULT   11
ID   R47208 standard; Protein; 147 AA.
AC   R47208;
DT   09-AUG-1994  (first entry)
DE   Human/murine IL-1 chimeric antibody VH.
KW   Probe; chimeric; recombinant; antibody; human; interleukin-1; IL-1;
KW   light; L; chain; constant; region; variable; mouse; anti-human;
KW   graft; CDR; complementarity determining region; heavy; H;
KW   inflammatory disease; arteriosclerosis; detection;
KW   diffused intravascular coagulation; leukemia.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1..19
FT                   /note= "Signal peptide"
FT   protein         20..147
FT                   /note= "Mature protein"
PN   WO9402627-A.
PD   03-FEB-1994.
PF   08-JUL-1993; J00941.
PR   16-JUL-1992; JP-189248.
PA   (SAKA ) OTSUKA PHARM CO LTD.
PI   Hirai Y,  Nishida T,  Omoto Y,  Owens RJ;
DR   WPI; 94-048885/06.
DR   N-PSDB; Q56069.
PT   Mouse/human chimeric antibody against human interleukin-1 - for
PT   treatment of diseases in which production of interleukin-1 is
PT   abnormal, and for diagnostic imaging of interleukin-1 production
PT   sites in vivo
PS   Claim 2; Fig 13; 58pp; Japanese.
CC   The sequences given in R47205-08 represent the light and heavy chain,
```

```
CC   variable and constant regions of a chimeric recombinant antibody
CC   against human interleukin-1 (IL-1).  The antibody has a light (L)
CC   chain in which the constant region is that of a human antibody and
CC   the variable region is from a mouse anti-human IL-1 antibody or is a
CC   mouse-human graft containing the CDR regions of mouse anti-human IL-1
CC   antibody, and a heavy (H) chain in which the constant region is that
CC   of a human antibody and the variable region is from a mouse anti-human
CC   IL-1 antibody or is a mouse-human graft containing the CDR regions of
CC   mouse anti-human IL-1 antibody,.  The chimeric antibody is used to
CC   treat diseases in which abnormal amounts of IL-1 are produced, eg.
CC   inflammatory disease, arteriosclerosis, diffused intravascular
CC   coagulation or leukemia.  It can also be labelled and used for
CC   diagnostic imaging of IL-1 producing sites in vivo.
SQ   Sequence   147 AA;

  Query Match           67.1%;  Score 55;  DB 9;  Length 147;
  Best Local Similarity  50.0%;  Pred. No. 1.12e+02;
  Matches     5;  Conservative    5;  Mismatches   0;  Indels   0;  Gaps   0;

Db    121 yfgssyaldy 130
          |:::|::|||
Qy      3 YLSTSFSLDY 12


RESULT   12
ID   R47211 standard; Protein; 147 AA.
AC   R47211;
DT   09-AUG-1994  (first entry)
DE   Chimeric human/mouse antibody H chain mature peptide.
KW   Probe; chimeric; recombinant; antibody; human; interleukin-1; IL-1;
KW   light; L; chain; constant; region; variable; mouse; anti-human; PCR;
KW   graft; CDR; complementarity determining region; heavy; H; detection;
KW   inflammatory disease; arteriosclerosis; polymerase chain reaction;
KW   diffused intravascular coagulation; leukemia; amplify.
OS   Chimeric - Homo sapiens.
OS   Chimeric - Mus musculus.
PN   WO9402627-A.
PD   03-FEB-1994.
PF   08-JUL-1993; J00941.
PR   16-JUL-1992; JP-189248.
PA   (SAKA ) OTSUKA PHARM CO LTD.
PI   Hirai Y,  Nishida T,  Omoto Y,  Owens RJ;
DR   WPI; 94-048885/06.
DR   N-PSDB; Q56076.
PT   Mouse/human chimeric antibody against human interleukin-1 - for
PT   treatment of diseases in which production of interleukin-1 is
PT   abnormal, and for diagnostic imaging of interleukin-1 production
PT   sites in vivo
PS   Disclosure; Fig 19; 58pp; Japanese.
CC   This sequence represents the heavy chain mature peptide of a chimeric
CC   recombinant antibody against human interleukin-1 (IL-1).  The chimeric
CC   antibody has a light (L) chain in which the constant region is that of
CC   a human antibody and the variable region is from a mouse anti-human
CC   IL-1 antibody or is a mouse-human graft containing the CDR regions of
CC   mouse anti-human IL-1 antibody, and a heavy (H) chain in which the
CC   constant region is that of a human antibody and the variable region is
CC   from a mouse anti-human IL-1 antibody or is a mouse-human graft
CC   containing the CDR regions of mouse anti-human IL-1 antibody.  The
CC   chimeric antibody is used to treat diseases in which abnormal amounts
CC   of IL-1 are produced, eg. inflammatory disease, arteriosclerosis,
CC   diffused intravascular coagulation or leukemia.  It can also be
CC   labelled and used for diagnostic imaging of IL-1 producing sites in
CC   vivo.
SQ   Sequence   147 AA;

  Query Match           67.1%;  Score 55;  DB 9;  Length 147;
  Best Local Similarity  50.0%;  Pred. No. 1.12e+02;
  Matches     5;  Conservative    5;  Mismatches   0;  Indels   0;  Gaps   0;

Db    121 yfgssyaldy 130
          |:::|::|||
Qy      3 YLSTSFSLDY 12


RESULT   13
ID   W47089 standard; Protein; 277 AA.
AC   W47089;
DT   20-JUL-1998  (first entry)
DE   Rat interleukin-1 beta converting enzyme-related protease.
KW   Interleukin-1 beta converting enzyme related protease;
KW   ICE related protease; rat; apoptosis; inhibitor; brain ischaemia;
KW   Alzheimer's disease; amyotrophic lateral sclerosis; head trauma;
KW   neurodegenerative disorder.
OS   Rattus sp.
PN   WO9803642-A1.
PD   29-JAN-1998.
PF   15-JUL-1997; U10999.
PR   22-JUL-1996; US-022049.
PA   (ELIL ) LILLY & CO ELI.
PI   Ni B, Paul SM, Wu X;
DR   WPI; 98-120767/11.
DR   N-PSDB; V13959.
PT   Rat interleukin-1 beta converting enzyme related protease - used for
PT   treating Alzheimer's disease, amyotrophic lateral sclerosis, and
PT   other neurodegenerative disorders
PS   Claim 1; Page 65; 76pp; English.
CC   This polypeptide comprises a novel rat interleukin-1 beta
CC   converting enzyme (ICE) related protease, overexpression of which
CC   may be involved in a neuronal death cascade.  The amino acid
CC   sequence was deduced from a cDNA clone (see V13959) isolated from a
CC   rat brain cDNA library.  ICE related protease is present in the
CC   central nervous system and is enriched in central neurons
CC   including pyramidal neurons and granule neurons of the hippocampus
CC   and cerebral cortex.  An expression vector capable of producing the
CC   ICE related protease in a host cell is claimed.  The protease can
CC   be used in a claimed method to identify compounds that inhibit
CC   apoptosis.  Such inhibitors can be used to treat Alzheimer's
CC   disease, amyotrophic lateral sclerosis (ALS), head trauma or other
CC   neurodegenerative disorders in humans (all claimed).  Antibodies
CC   raised against the protease may be usef dor diagnosis of these
CC   disease states, or in their treatment.
SQ   Sequence   277 AA;

  Query Match           65.9%;  Score 54;  DB 30;  Length 277;
  Best Local Similarity  50.0%;  Pred. No. 1.39e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     32 yldssykmdy 41
          || :|: :||
Qy      3 YLSTSFSLDY 12


RESULT   14
ID   W64560 standard; Protein; 783 AA.
AC   W64560;
DT   07-DEC-1998  (first entry)
DE   Candida albicans oligopeptide transporter.
KW   Oligopeptide transporter; OPT1 gene; fungicide; antifungal;
KW   anticandidal.
OS   Candida albicans strain 1006.
FH   Key             Location/Qualifiers
FT   Misc_difference 47
FT                   /note= "encoded by CAA"
FT   Misc_difference 152
FT                   /note= "encoded by ATT"
FT   Misc_difference 403
FT                   /note= "encoded by TTC"
FT   Misc_difference 436
FT                   /note= "encoded by GAG"
FT   Misc_difference 447
FT                   /note= "encoded by CAA"
FT   Misc_difference 469
FT                   /note= "encoded by TCC"
FT   Misc_difference 474
FT                   /note= "encoded by GTT"
FT   Misc_difference 515
```

FT /note= "encoded by AAT"
FT Misc_difference 659
FT /note= "encoded by TGG"
FT Misc_difference 733
FT /note= "encoded by TGC"
PN WO9834950-A1.
PD 13-AUG-1998.
PF 06-FEB-1998; U02332.
PR 07-FEB-1997; US-037859.
PA (UYTE-) UNIV TENNESSEE RES CORP.
PI Becker JM, Lubkowitz MA;
DR WPI; 98-447166/38.
DR N-PSDB; V49601.
PT New oligopeptide membrane transporter from Candida albicans and
PT related nucleic acid - used to deliver therapeutic agents,
PT especially antifungals, to target cells and to identify plant
PT transporters
PS Claim 6; Fig 3a-c; 44pp; English.
CC This is the Candida albicans oligopeptide transporter, OPT1. Its
CC amino acid sequence was deduced from an isolated genomic clone (see
CC V49601) obtained through heterologous expression in Saccharomyces
CC cerevisiae di-/tripeptide transport mutant PB1X-9B. When
CC transformed with a plasmid harboring OPT1, mutant PX1X-9B, which
CC does not express tetra-/pentapeptide transport activity under the
CC conditions used, was conferred with an oligopeptide transport
CC phenotype as indicated by growth on Lys-Leu-Leu-Gly, sensitivity to
CC toxic tetra- and pentapeptides, and an increase in the initial
CC uptake rate of the radiolabeled tetrapeptide Lys-Leu-Gly-(3H)Leu.
CC The deduced protein product contains 12 hydrophobic regions,
CC suggestive of a membrane transport protein. The oligopeptide
CC transporter facilitates the uptake of normally non-permeable drugs,
CC and is especially used to target antifungals, specifically
CC anti-Candida drugs, which are conjugated, linked or attached to the
CC peptide being transported. It can also be used to identify plant
CC oligopeptide transporter genes, making it possible to use
CC oligopeptides as herbicides or growth stimulants.
SQ Sequence 783 AA;

Query Match 64.6%; Score 53; DB 35; Length 783;
Best Local Similarity 50.0%; Pred. No. 1.71e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

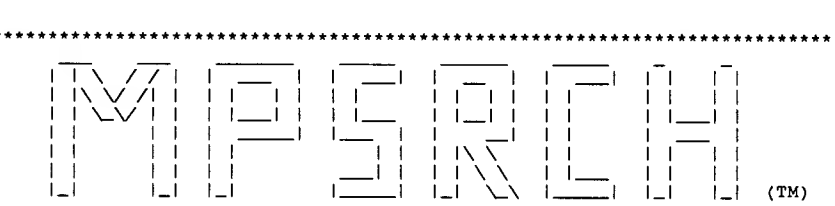
Db 406 flsttfaisy 415
:|||:|:: |
Qy 3 YLSTSFSLDY 12

RESULT 15
R83040 standard; peptide; 306 AA.
R83040;
DT 03-APR-1996 (first entry)
DE Capsular polysaccharide of Streptococcus pneumoniae (CpsU).
KW Capsular polysaccharide; cps; peptide; flanking region; detection;
KW serotype; diagnosis; prevention; Streptococcus pneumoniae.
OS Streptococcus pneumoniae.
PN WO9531548-A1.
PD 23-NOV-1995.
PF 16-MAY-1995; U06119.
PR 16-MAY-1994; US-243546.
PA (UABR-) UAB RES FOUND.
PI Dillard JP, Yother J;
DR WPI; 96-010934/01.
DR N-PSDB; T05848.
PT New Streptococcus pneumoniae capsular polysaccharide genes - used
PT for detection, serotyping and for diagnosis and prevention of S.
PT pneumoniae infection
PS Disclosure; Page 178-180; 226pp; English.
CC Sequences encoding the 5' flanking region of the capsular
CC polysaccharide gene (cps) of Streptococcus pneumoniae and which are
CC of sufficient length to allow hybridisation under standard
CC hybridisation conditions to a S.pneumoniae cps gene flanking region
CC may be used in methods to detect and serotype S.pneumoniae. They
CC may also be used for the diagnosis and prevention of S. pneumoniae
CC infection.
SQ Sequence 306 AA;

Query Match 63.4%; Score 52; DB 15; Length 306;
Best Local Similarity 40.0%; Pred. No. 2.11e+02;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 67 yfdstfeley 76
|: ::| |:|
Qy 3 YLSTSFSLDY 12

Search completed: Thu Sep 2 12:38:34 1999
Job time : 20 secs.



MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:39:53 1999; MasPar time 4.71 Seconds
138.937 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-31
Description: (1-12) from US08599226.pep
Perfect Score: 82
Sequence: 1 ASYLSTSFSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.125; Variance 38.161; scale 0.658

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	63	76.8	372	5	P91143	SIMILAR TO ACETYLTRANS	6.49e-01
2	61	74.4	444	5	P91141	SIMILAR TO ACETYLTRANS	1.49e+00
3	59	72.0	472	2	O86563	TRANSMEMBRANE TRANSPOR	3.39e+00
4	59	72.0	1367	5	Q20120	F44G4.8 PROTEIN (EC 3.	3.39e+00
5	56	68.3	315	2	Q55611	PROTEIN-EXPORT MEMBRAN	1.12e+01
6	55	67.1	306	6	O77768	HETEROGENEOUS NUCLEAR	1.65e+01
7	55	67.1	357	2	O05237	HYPOTHETICAL 39.5 KD P	1.65e+01
8	55	67.1	360	8	O47575	CYTOCHROME B.	1.65e+01
9	55	67.1	603	8	O79437	NADH DEHYDROGENASE SUB	1.65e+01
10	54	65.9	84	3	Q04203	HYPOTHETICAL 9.9 KD PR	2.43e+01
11	54	65.9	193	3	Q12064	D1554.	2.43e+01
12	54	65.9	262	2	O26039	CONSERVED HYPOTHETICAL	2.43e+01
13	54	65.9	316	2	O65998	ORF36.	2.43e+01
14	53	64.6	389	5	Q20272	COSMID F41C6.	3.56e+01
15	53	64.6	783	3	O14411	OPT1P.	3.56e+01
16	52	63.4	248	1	O59270	248AA LONG HYPOTHETICA	5.18e+01
17	52	63.4	253	2	P96151	DNA FOR AMINOPEPTIDASE	5.18e+01
18	52	63.4	265	10	O24226	CHLOROPHYLL A/B BINDIN	5.18e+01
19	52	63.4	265	10	O64442	LIGHT HARVESTING CHLOR	5.18e+01
20	52	63.4	336	2	Q47450	PAPG PROTEIN.	5.18e+01
21	52	63.4	336	2	Q47445	PAPG PROTEIN.	5.18e+01
22	52	63.4	336	2	Q47444	PAPE.	5.18e+01
23	52	63.4	337	2	Q47448	PAPG PROTEIN.	5.18e+01
24	52	63.4	405	5	O17660	C41G6.7 PROTEIN.	5.18e+01
25	52	63.4	450	2	O24913	CONSERVED HYPOTHETICAL	5.18e+01
26	52	63.4	458	2	P73215	DNA POLYMERASE III ALP	5.18e+01
27	52	63.4	551	11	O88446	POTENTIAL-SENSITIVE PO	5.18e+01
28	52	63.4	556	4	O75751	EXTRANEURONAL MONOAMIN	5.18e+01
29	52	63.4	588	2	P96828	HYPOTHETICAL 61.4 KD P	5.18e+01
30	52	63.4	595	10	O49675	NEOXANTHIN CLEAVAGE EN	5.18e+01
31	52	63.4	649	2	O84307	ATP SYNTHASE SUBUNIT I	5.18e+01
32	52	63.4	722	10	O04514	SIMILAR TO S. POMBE IS	5.18e+01
33	51	62.2	118	10	O81621	F8M12.8 PROTEIN.	7.50e+01
34	51	62.2	153	5	O45791	T19C9.6 PROTEIN.	7.50e+01
35	51	62.2	189	10	O22237	HYPOTHETICAL 22.1 KD P	7.50e+01
36	51	62.2	244	2	O07422	HYPOTHETICAL 25.9 KD P	7.50e+01
37	51	62.2	341	5	Q18049	SIMILARITY TO C. ELEGA	7.50e+01
38	51	62.2	422	4	O75462	CYTOKINE-LIKE FACTOR-1	7.50e+01
39	51	62.2	428	5	Q17394	TRANSMEMBRANE PROTEIN.	7.50e+01
40	51	62.2	454	2	O51495	XYLULOKINASE (XYLB).	7.50e+01
41	51	62.2	582	5	Q22774	HYPOTHETICAL PROTEIN T	7.50e+01
42	51	62.2	2014	5	O01703	T25C12.3 PROTEIN.	7.50e+01
43	51	62.2	3391	14	P89313	POLYPROTEIN.	7.50e+01
44	51	62.2	3391	14	P89314	POLYPROTEIN.	7.50e+01
45	51	62.2	3391	14	P89315	POLYPROTEIN.	7.50e+01

ALIGNMENTS

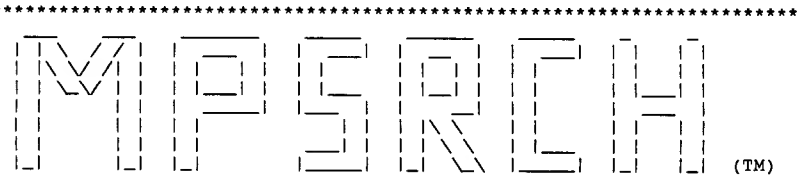
```
RESULT    1
ID   P91143      PRELIMINARY;      PRT;   372 AA.
AC   P91143;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO ACETYLTRANSFERASES.
GN   C37H5.2.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DAVIDSON S., GILLAM B.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U88315; G1825777; -.
DR   PFAM; PF00561; abhydrolase; 1.
KW   TRANSFERASE.
SQ   SEQUENCE   372 AA;  42139 MW;  5214F159 CRC32;

  Query Match           76.8%;  Score 63;  DB 5;  Length 372;
```

```
  Best Local Similarity  58.3%;  Pred. No. 6.49e-01;
  Matches     7;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    168 GGYLSTSYALKY 179
          ::|||||::| |
Qy      1 ASYLSTSFSLDY 12


RESULT    2
ID   P91141      PRELIMINARY;      PRT;   444 AA.
AC   P91141;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO ACETYLTRANSFERASES.
GN   C37H5.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DAVIDSON S., GILLAM B.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U88315; G1825778; -.
DR   PFAM; PF00561; abhydrolase; 1.
KW   TRANSFERASE.
SQ   SEQUENCE   444 AA;  50487 MW;  F706E46B CRC32;

  Query Match           74.4%;  Score 61;  DB 5;  Length 444;
  Best Local Similarity  50.0%;  Pred. No. 1.49e+00;
  Matches     6;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    240 GGYLATSYALKY 251
          ::||:||::| |
Qy      1 ASYLSTSFSLDY 12


RESULT    3
ID   O86563      PRELIMINARY;      PRT;   472 AA.
AC   O86563;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRANSMEMBRANE TRANSPORT PROTEIN.
GN   SC2A11.02C.
OS   STREPTOMYCES COELICOLOR.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; STREPTOMYCINEAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   MURPHY L., HARRIS D.;
RL   SUBMITTED (AUG-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   SUBMITTED (AUG-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE; 97000351.
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   MOL. MICROBIOL. 21:77-96(1996).
DR   EMBL; AL031184; E1314374; -.
KW   TRANSMEMBRANE.
SQ   SEQUENCE   472 AA;  50203 MW;  66C4596E CRC32;

  Query Match           72.0%;  Score 59;  DB 2;  Length 472;
  Best Local Similarity  70.0%;  Pred. No. 3.39e+00;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    272 YLMTTFSLSY 281
          || |:||| |
Qy      3 YLSTSFSLDY 12


RESULT    4
ID   Q20120      PRELIMINARY;      PRT;  1367 AA.
AC   Q20120;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   F44G4.8 PROTEIN (EC 3.1.3.48).
GN   F44G4.8.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   THOMAS K.;
RL   SUBMITTED (SEP-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z54218; E1346536; -.
DR   EMBL; Z49910; E1346536; JOINED.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW   HYDROLASE.
SQ   SEQUENCE   1367 AA;  152872 MW;  F696EAD5 CRC32;

  Query Match           72.0%;  Score 59;  DB 5;  Length 1367;
  Best Local Similarity  63.6%;  Pred. No. 3.39e+00;
  Matches     7;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    160 SYLQDSFSIEY 170
          |||  |||::|
Qy      2 SYLSTSFSLDY 12


RESULT    5
ID   Q55611      PRELIMINARY;      PRT;   315 AA.
AC   Q55611;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   PROTEIN-EXPORT MEMBRANE PROTEIN SECF.
GN   SECF.
OS   SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC   BACTERIA; CYANOBACTERIA; CHROOCOCCALES; SYNECHOCYSTIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
```

```
RA   TABATA S.;
RL   SUBMITTED (AUG-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 96127529.
RA   KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA   SUGIURA M., TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. sequence features in the 1Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA RES. 2:153-166(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. PCC6803. II. Sequence determination of the entire
RT   genome and assignment of potential protein-coding regions.";
RL   DNA RES. 3:109-136(1996).
DR   EMBL; D64000; D1010770; -.
SQ   SEQUENCE   315 AA;  34667 MW;  4C20ED73 CRC32;

  Query Match           68.3%;  Score 56;  DB 2;  Length 315;
  Best Local Similarity  60.0%;  Pred. No. 1.12e+01;
  Matches     6;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Db    155 YLTIRFQLDY 164
          ||:  | |||
Qy      3 YLSTSFSLDY 12


RESULT    6
ID   O77768      PRELIMINARY;      PRT;   306 AA.
AC   O77768;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN C.
GN   HNRNPC.
OS   ORYCTOLAGUS CUNICULUS (RABBIT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ENGLAND WHITE;
RX   MEDLINE; 98438739.
RA   JIANG W., GUO X., BHAVANANDAN V.P.;
RT   "Four distinct regions in the auxiliary domain of heterogeneous
RT   nuclear ribonucleoprotein C-related proteins.";
RL   BIOCHIM. BIOPHYS. ACTA 1399:229-233(1998).
DR   EMBL; AF061582; G3660678; -.
KW   NUCLEOPROTEIN; RIBONUCLEOPROTEIN.
SQ   SEQUENCE   306 AA;  33684 MW;  FBB28492 CRC32;

  Query Match           67.1%;  Score 55;  DB 6;  Length 306;
  Best Local Similarity  77.8%;  Pred. No. 1.65e+01;
  Matches     7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    118 LSSSFDLDY 126
          ||:|| |||
Qy      4 LSTSFSLDY 12


RESULT    7
ID   O05237      PRELIMINARY;      PRT;   357 AA.
AC   O05237;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 39.5 KD PROTEIN.
GN   YUGH.
OS   BACILLUS SUBTILIS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   OUDEGA B., KONINGSTEYN G.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE OF 1-214 FROM N.A.
RC   STRAIN=168;
RA   DANCHIN A.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   OUDEGA B.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA   AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA   BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA   BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA   CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA   DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA   ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA   FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA   GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA   GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA   HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA   JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA   KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA   KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA   LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA   MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA   NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA   PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA   PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA   RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA   SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA   SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
RA   SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA   TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA   TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA   VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA   WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA   YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   NATURE 390:249-256(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
DR   EMBL; Z93934; E311527; -.
DR   EMBL; Z99120; E1184218; -.
DR   PROSITE; PS00105; AA_TRANSFER_CLASS_1; 1.
DR   PFAM; PF00155; aminotran_1; 1.
KW   HYPOTHETICAL PROTEIN; PYRIDOXAL PHOSPHATE.
SQ   SEQUENCE   357 AA;  39461 MW;  FD6B02F9 CRC32;

  Query Match           67.1%;  Score 55;  DB 2;  Length 357;
```

```
  Best Local Similarity  50.0%;  Pred. No. 1.65e+01;
  Matches     6;  Conservative    1;  Mismatches   5;  Indels   0;  Gaps   0;

Db     75 SRYLSNRFDLSY 86
          : |||  | | |
Qy      1 ASYLSTSFSLDY 12


RESULT    8
ID   O47575      PRELIMINARY;      PRT;   360 AA.
AC   O47575;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   CYTOCHROME B.
OS   ONCHOCERCA VOLVULUS.
OG   MITOCHONDRION.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; SPIRURIA; SPIRURIDA;
OC   FILARIOIDEA; ONCHOCERCIDAE; ONCHOCERCA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FOREST;
RA   KEDDIE E.M., UNNASCH T.R.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q + 2
CC       FERROCYTOCHROME C.
CC   -!- COFACTOR: TWO HEME GROUPS
CC       (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
CC       (BY SIMILARITY).
DR   EMBL; AF015193; G2735939; -.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
KW   MITOCHONDRION; ELECTRON TRANSPORT; RESPIRATORY CHAIN; TRANSMEMBRANE;
KW   HEME.
SQ   SEQUENCE   360 AA;  42544 MW;  E80DD32E CRC32;

  Query Match           67.1%;  Score 55;  DB 8;  Length 360;
  Best Local Similarity  50.0%;  Pred. No. 1.65e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db      9 FLPASFTLSY 18
          :|::||:| |
Qy      3 YLSTSFSLDY 12


RESULT    9
ID   O79437      PRELIMINARY;      PRT;   603 AA.
AC   O79437;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NADH DEHYDROGENASE SUBUNIT 5.
GN   NADH5.
OS   ORYCTOLAGUS CUNICULUS (RABBIT).
OG   MITOCHONDRION.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98317530.
RA   GISSI C., GULLBERG A., ARNASON U.;
RT   "The complete mitochondrial DNA sequence of the rabbit, Oryctolagus
RT   cuniculus.";
RL   GENOMICS 50:161-169(1998).
DR   EMBL; AJ001588; E1310026; -.
KW   MITOCHONDRION.
SQ   SEQUENCE   603 AA;  67389 MW;  7A09A8C3 CRC32;

  Query Match           67.1%;  Score 55;  DB 8;  Length 603;
  Best Local Similarity  77.8%;  Pred. No. 1.65e+01;
  Matches     7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db     76 LTTSFKLDY 84
          |:||| |||


Qy      4 LSTSFSLDY 12


RESULT   10
ID   Q04203      PRELIMINARY;      PRT;    84 AA.
AC   Q04203;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 9.9 KD PROTEIN.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB972;
RA   BOWMAN S.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB972;
RA   BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z49919; G887589; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   84 AA;  9904 MW;  C2D238D3 CRC32;

  Query Match           65.9%;  Score 54;  DB 3;  Length 84;
  Best Local Similarity  60.0%;  Pred. No. 2.43e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db      6 YLNYAFSLAY 15
          ||: :||| |
Qy      3 YLSTSFSLDY 12


RESULT   11
ID   Q12064      PRELIMINARY;      PRT;   193 AA.
AC   Q12064;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE   D1554.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DELAVEAU T., BLUGEON C., JACQ C., PEREA J.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   PEREA J., BLUGEON C., DELAVEAU T., JACQ C.;
RL   SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; X97751; E242701; -.
DR   EMBL; Z74199; E253062; -.
SQ   SEQUENCE   193 AA;  20893 MW;  37629BA8 CRC32;

  Query Match           65.9%;  Score 54;  DB 3;  Length 193;
  Best Local Similarity  40.0%;  Pred. No. 2.43e+01;
  Matches     4;  Conservative    6;  Mismatches   0;  Indels   0;  Gaps   0;

Db    149 FLATAFGLNF 158
          :|:|:|:|::
Qy      3 YLSTSFSLDY 12


RESULT   12
ID   O26039      PRELIMINARY;      PRT;   262 AA.
```

```
AC   O26039;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN.
GN   HP1509.
OS   HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC   BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC   HELICOBACTER.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695;
RX   MEDLINE; 97394467.
RA   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori[published erratum appears in Nature 1997 Sep
RT   25;389(6649):412].";
RL   NATURE 388:539-547(1997).
DR   EMBL; AE000649; G2314690; -.
DR   TIGR; HP1509; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   262 AA;  28835 MW;  5CE61FE3 CRC32;

  Query Match           65.9%;  Score 54;  DB 2;  Length 262;
  Best Local Similarity  58.3%;  Pred. No. 2.43e+01;
  Matches    7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db    125 AVFLSKLFGLDY 136
          | :||  |:|||
Qy      1 ASYLSTSFSLDY 12


RESULT   13
ID   O65998      PRELIMINARY;      PRT;   316 AA.
AC   O65998;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   ORF36.
OS   CLOSTRIDIUM ACETOBUTYLICUM.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC   CLOSTRIDIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 792;
RA   BEHRENS S.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U58131; G3025464; -.
SQ   SEQUENCE   316 AA;  36225 MW;  720478E4 CRC32;

  Query Match           65.9%;  Score 54;  DB 2;  Length 316;
  Best Local Similarity  41.7%;  Pred. No. 2.43e+01;
  Matches    5;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db    207 PSYLPSGYKFDY 218
          :|||:::: :||
Qy      1 ASYLSTSFSLDY 12


RESULT   14
ID   Q20272      PRELIMINARY;      PRT;   389 AA.
AC   Q20272;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   COSMID F41C6.
GN   F41C6.7.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   GEISEL C.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
RL   SUBMITTED (OCT-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U39745; G1049476; -.
SQ   SEQUENCE   389 AA;  44253 MW;  FFE0AFC6 CRC32;

  Query Match           64.6%;  Score 53;  DB 5;  Length 389;
  Best Local Similarity  70.0%;  Pred. No. 3.56e+01;
  Matches    7;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    157 AAYLSSSFSI 166
          |:|||:|||:
Qy      1 ASYLSTSFSL 10


RESULT   15
ID   O14411      PRELIMINARY;      PRT;   783 AA.
AC   O14411;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   OPT1P.
GN   OPT1.
OS   CANDIDA ALBICANS (YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   CANDIDACEAE; CANDIDA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1006;
RX   MEDLINE; 97195785.
RA   LUBKOWITZ M.A., HAUSER L., BRESLAV M., NAIDER F., BECKER J.M.;
RT   "An oligopeptide transport gene from Candida albicans.";
RL   MICROBIOLOGY 143:0-0(0).
DR   EMBL; U60973; G2367386; -.
SQ   SEQUENCE   783 AA;  88406 MW;  DC99C92E CRC32;

  Query Match           64.6%;  Score 53;  DB 3;  Length 783;
  Best Local Similarity  50.0%;  Pred. No. 3.56e+01;
  Matches    5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    406 FLSTTFAISY 415
          :|||:|:: |
Qy      3 YLSTSFSLDY 12


Search completed: Thu Sep  2 12:40:23 1999
Job time : 30 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

[redacted] on: Thu Sep 2 12:39:25 1999; MasPar time 2.20 Seconds
154.427 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-31
Description: (1-12) from US08599226.pep
Perfect Score: 82
Sequence: 1 ASYLSTSFSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 26.265; Variance 33.740; scale 0.778

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	63	76.8	272	1	YIGL_HAEIN	HYPOTHETICAL PROTEIN H	1.04e-01
2	62	75.6	1294	1	YA3B_SCHPO	HYPOTHETICAL 149.2 KD	1.67e-01
3	57	69.5	296	1	CYCG_RHOSH	DIHEME CYTOCHROME C-TY	1.69e+00
4	57	69.5	300	1	GP40_HUMAN	PUTATIVE G PROTEIN-COU	1.69e+00
5	56	68.3	575	1	CNGX_RAT	CYCLIC-NUCLEOTIDE-GATE	2.64e+00
6	55	67.1	303	1	ROC_HUMAN	HETEROGENEOUS NUCLEAR	4.10e+00
7	55	67.1	438	1	SHIA_ECOLI	SHIKIMATE TRANSPORTER.	4.10e+00
8	55	67.1	445	1	ACSC_MOOTH	CORRINOID/IRON-SULFUR	4.10e+00
9	54	65.9	277	1	ICE3_CRILO	APOPAIN PRECURSOR (EC	6.33e+00
10	54	65.9	277	1	ICE3_RAT	APOPAIN PRECURSOR (EC	6.33e+00
11	54	65.9	277	1	ICE3_MOUSE	APOPAIN PRECURSOR (EC	6.33e+00
12	54	65.9	611	1	YCR3_YEAST	HYPOTHETICAL 69.2 KD P	6.33e+00
13	53	64.6	175	1	RL6B_YEAST	60S RIBOSOMAL PROTEIN	9.73e+00
14	53	64.6	175	1	RL6A_YEAST	60S RIBOSOMAL PROTEIN	9.73e+00
15	53	64.6	324	1	GSHB_ANASP	GLUTATHIONE SYNTHETASE	9.73e+00
16	53	64.6	738	1	ST11_YEAST	SERINE/THREONINE-PROTE	9.73e+00
17	52	63.4	122	1	YG27_YEAST	HYPOTHETICAL 14.4 KD P	1.49e+01
18	52	63.4	180	1	APT_MASHI	ADENINE PHOSPHORIBOSYL	1.49e+01
19	52	63.4	180	1	APT_STOLO	ADENINE PHOSPHORIBOSYL	1.49e+01
20	52	63.4	227	1	PGSA_MYCPN	CDP-DIACYLGLYCEROL--GL	1.49e+01
21	52	63.4	306	1	CAPC_STRPN	UTP--GLUCOSE-1-PHOSPHA	1.49e+01
22	52	63.4	428	1	YE63_SCHPO	HYPOTHETICAL 48.7 KD P	1.49e+01
23	52	63.4	475	1	YM61_YEAST	HYPOTHETICAL 55.3 KD P	1.49e+01
24	51	62.2	53	1	VG87_BPML5	GENE 87 PROTEIN (GP87)	2.25e+01
25	51	62.2	150	1	YB9L_YEAST	HYPOTHETICAL 16.0 KD P	2.25e+01
26	51	62.2	265	1	CB23_POLMU	CHLOROPHYLL A-B BINDIN	2.25e+01
27	51	62.2	552	1	CYHR_CANMA	CYCLOHEXIMIDE RESISTAN	2.25e+01
28	51	62.2	579	1	YR47_CAEEL	HYPOTHETICAL 66.0 KD P	2.25e+01
29	51	62.2	715	1	LCNC_LACLA	LACTOCOCCIN A TRANSPOR	2.25e+01
30	51	62.2	799	1	AFSK_STRCO	SERINE/THREONINE PROTE	2.25e+01
31	51	62.2	807	1	AFSK_STRGR	SERINE/THREONINE PROTE	2.25e+01
32	51	62.2	817	1	TGLK_HUMAN	PROTEIN-GLUTAMINE GAMM	2.25e+01
33	51	62.2	824	1	TGLK_RAT	PROTEIN-GLUTAMINE GAMM	2.25e+01
34	51	62.2	836	1	TGLK_RABIT	PROTEIN-GLUTAMINE GAMM	2.25e+01
35	51	62.2	3396	1	POLG_DEN1S	GENOME POLYPROTEIN [CO	2.25e+01
36	50	61.0	179	1	APT_HELPY	ADENINE PHOSPHORIBOSYL	3.39e+01
37	50	61.0	241	1	ATP6_RHORU	ATP SYNTHASE A CHAIN (	3.39e+01
38	50	61.0	267	1	CB23_PETSP	CHLOROPHYLL A-B BINDIN	3.39e+01
39	50	61.0	406	1	GUN1_RUMAL	ENDOGLUCANASE I PRECUR	3.39e+01
40	50	61.0	409	1	GUNB_RUMAL	ENDOGLUCANASE B PRECUR	3.39e+01
41	50	61.0	577	1	SYR_HAEIN	ARGINYL-TRNA SYNTHETAS	3.39e+01
42	50	61.0	700	1	CAOQ_RAT	PRISTANOYL-COA OXIDASE	3.39e+01
43	50	61.0	835	1	FASD_ECOLI	OUTER MEMBRANE USHER P	3.39e+01
44	50	61.0	844	1	MCEL_VARV	MRNA CAPPING ENZYME, L	3.39e+01
45	50	61.0	844	1	MCEL_VACCC	MRNA CAPPING ENZYME, L	3.39e+01

ALIGNMENTS

RESULT 1
ID YIGL_HAEIN STANDARD; PRT; 272 AA.
AC P44771;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN HI0597.
GN HI0597.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL SCIENCE 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO THE COF/YBHA/YIDA/YIGL (E.COLI) / YCSE/YXEH
CC (B.SUBTILIS) FAMILY. STRONG, TO E.COLI YIGL.
CC ---

CC ---
DR EMBL; U32741; G1573586; -.
DR TIGR; HI0597; -.
DR PROSITE; PS01228; COF_1; 1.
DR PROSITE; PS01229; COF_2; 1.
DR PFAM; PF00592; DUF3; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 272 AA; 30523 MW; 7F53B65C CRC32;

Query Match 76.8%; Score 63; DB 1; Length 272;

```
  Best Local Similarity  66.7%;  Pred. No. 1.04e-01;
  Matches     8;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Db    261 ARYLRTQFGLDY 272
          | || | |:|||
Qy      1 ASYLSTSFSLDY 12


RESULT    2
ID   YA3B_SCHPO     STANDARD;      PRT;  1294 AA.
AC   Q09716;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 149.2 KD PROTEIN C18B11.11 IN CHROMOSOME I.
GN   SPAC18B11.11 OR SPAC1F5.01.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (AUG-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z50728; G929897; -.
DR   EMBL; Z68136; E211997; -.
KW   HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT   TRANSMEM     205     225       POTENTIAL.
FT   TRANSMEM     277     297       POTENTIAL.
FT   TRANSMEM     553     573       POTENTIAL.
FT   TRANSMEM     888     908       POTENTIAL.
SQ   SEQUENCE   1294 AA;  149192 MW;  7615C3B4 CRC32;

  Query Match           75.6%;  Score 62;  DB 1;  Length 1294;
  Best Local Similarity  72.7%;  Pred. No. 1.67e-01;
  Matches     8;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    456 SYFDTSFSLDF 466
          ||: ||||||:
Qy      2 SYLSTSFSLDY 12


RESULT    3
ID   CYCG_RHOSH     STANDARD;      PRT;   296 AA.
AC   Q53143;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   DIHEME CYTOCHROME C-TYPE.
GN   CYCG.
OS   RHODOBACTER SPHAEROIDES (RHODOPSEUDOMONAS SPHAEROIDES).
OC   BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;
OC   RHODOBACTER.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2.4.1;
RX   MEDLINE; 95362655.
RA   FLORY J.E., DONOHUE T.J.;
RT   "Organization and expression of the Rhodobacter sphaeroides cycFG
RT   operon.";
RL   J. BACTERIOL. 177:4311-4320(1995).
CC   -!- FUNCTION: DIHEME C-TYPE CYTOCHROME, THAT IS PARTICULARLY EXPRESSED
CC       WHEN CELLS GENERATE ENERGY VIA AEROBIC RESPIRATION.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC   -!- PTM: BINDS TWO HEME GROUPS PER MOLECULE (POTENTIAL).
CC   -!- SIMILARITY: TO ACETOBACTER ALCOHOL DEHYDROGENASE CYTOCHROME C
CC       SUBUNIT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L36880; G557745; -.
DR   PROSITE; PS00190; CYTOCHROME_C; 2.
DR   PFAM; PF00034; cytochrome_c; 1.
KW   ELECTRON TRANSPORT; HEME; MEMBRANE.
FT   BINDING      52      52       HEME 1 (COVALENT) (BY SIMILARITY).
FT   BINDING      55      55       HEME 1 (COVALENT) (BY SIMILARITY).
FT   METAL        56      56       IRON 1 (HEME AXIAL LIGAND) (BY
FT                                 SIMILARITY).
FT   BINDING     202     202       HEME 2 (COVALENT) (BY SIMILARITY).
FT   BINDING     205     205       HEME 2 (COVALENT) (BY SIMILARITY).
FT   METAL       206     206       IRON 2 (HEME AXIAL LIGAND) (BY
FT                                 SIMILARITY).
SQ   SEQUENCE   296 AA;  31727 MW;  440EB356 CRC32;

  Query Match           69.5%;  Score 57;  DB 1;  Length 296;
  Best Local Similarity  58.3%;  Pred. No. 1.69e+00;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    249 AEYLSSGFTPDY 260
          | |||::|: ||
Qy      1 ASYLSTSFSLDY 12


RESULT    4
ID   GP40_HUMAN     STANDARD;      PRT;   300 AA.
AC   O14842;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PUTATIVE G PROTEIN-COUPLED RECEPTOR GPR40.
GN   GPR40.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98008875.
RA   SAWZDARGO M., GEORGE S.R., NGUYEN T., XU S., KOLAKOWSKI L.F. JR.,
RA   O'DOWD B.F.;
RT   "A cluster of four novel human G protein-coupled receptor genes
RT   occurring in close proximity to CD22 gene on chromosome 19q13.1.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 239:543-547(1997).
CC   -!- FUNCTION: ORPHAN RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF024687; G2612946; -.
DR   GCRDB; GCR_2538; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
DR   PFAM; PF00001; 7tm_1; 2.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE.
```

```
FT   DOMAIN        1     10       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     11     31       1 (POTENTIAL).
FT   DOMAIN       32     39       CYTOPLASMIC.
FT   TRANSMEM     40     59       2 (POTENTIAL).
FT   DOMAIN       60     71       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     72     92       3 (POTENTIAL).
FT   DOMAIN       93    123       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    124    144       4 (POTENTIAL).
FT   DOMAIN      145    178       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    179    197       5 (POTENTIAL).
FT   DOMAIN      198    221       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    222    242       6 (POTENTIAL).
FT   DOMAIN      243    258       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    259    279       7 (POTENTIAL).
FT   DOMAIN      280    300       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    155    155       POTENTIAL.
SQ   SEQUENCE   300 AA;  31457 MW;  24383E01 CRC32;

  Query Match           69.5%;  Score 57;  DB 1;  Length 300;
  Best Local Similarity 50.0%;  Pred. No. 1.69e+00;
  Matches    5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    105 YLGAAFPLGY 114
          ||:::|:| |
Qy      3 YLSTSFSLDY 12


RESULT    5
ID   CNGX_RAT       STANDARD;      PRT;   575 AA.
AC   Q64359;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CYCLIC-NUCLEOTIDE-GATED OLFACTORY CHANNEL OCNC2 SUBUNIT.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE; 94377458.
RA   BRADLEY J., LI J., DAVIDSON N., LESTER H.A., ZINN K.;
RT   "Heteromeric olfactory cyclic nucleotide-gated channels: a subunit
RT   that confers increased sensitivity to cAMP.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 91:8890-8894(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OLFACTORY NEUROEPITHELIUM;
     MEDLINE; 95000663.
     LIMAN E.R., BUCK L.B.;
RT   "A second subunit of the olfactory cyclic nucleotide-gated channel
RT   confers high sensitivity to cAMP.";
RL   NEURON 13:611-621(1994).
RN   [3]
RP   SEQUENCE OF 7-35 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RA   BRADLEY J., ZHANG Y., BAKIN R., LESTER H.A., RONNETT G., ZINN K.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: ODORANT SIGNAL TRANSDUCTION IS PROBABLY MEDIATED
CC       BY A G-PROTEIN COUPLED CASCADE USING CAMP AS SECOND MESSENGER.
CC       THE OLFACTORY CHANNEL CAN BE SHOWN TO BE ACTIVATED BY CYCLIC
CC       NUCLEOTIDES WHICH LEADS TO A DEPOLARIZATION OF OLFACTORY
CC       SENSORY NEURONS.
CC   -!- SUBUNIT: HETEROOLIGOMER OF OCNC1 AND OCNC2 SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: OLFACTORY NEURONS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12623; G538129; -.
DR   EMBL; U12425; G548084; -.
DR   EMBL; U76219; G1753128; -.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
DR   PFAM; PF00027; cNMP_binding; 1.
DR   PFAM; PF00914; CNG_membrane; 1.
KW   IONIC CHANNEL; ION TRANSPORT; CAMP-BINDING; TRANSMEMBRANE;
KW   MULTIGENE FAMILY; OLFACTION.
FT   DOMAIN        1     33       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     34     54       H1 (POTENTIAL).
FT   DOMAIN       55     65       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     66     86       H2 (POTENTIAL).
FT   DOMAIN       87    112       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    113    133       H3 (POTENTIAL).
FT   DOMAIN      134    168       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    169    189       H4 (POTENTIAL).
FT   DOMAIN      190    244       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    245    265       H5 (POTENTIAL).
FT   DOMAIN      266    400       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    401    421       H6 (POTENTIAL).
FT   DOMAIN      422    575       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     356    500       CAMP (BY SIMILARITY).
FT   BINDING     415    415       CAMP (POTENTIAL).
FT   BINDING     430    430       CAMP (POTENTIAL).
SQ   SEQUENCE   575 AA;  65674 MW;  C545A708 CRC32;

  Query Match           68.3%;  Score 56;  DB 1;  Length 575;
  Best Local Similarity 63.6%;  Pred. No. 2.64e+00;
  Matches    7;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Db     64 SYLVAWFVLDY 74
          ||| : | |||
Qy      2 SYLSTSFSLDY 12


RESULT    6
ID   ROC_HUMAN      STANDARD;      PRT;   303 AA.
AC   P07910; P22628;
DT   01-AUG-1988 (REL. 08, CREATED)
DT   01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEINS C1/C2 (HNRNP C1 AND HNRNP
DE   C2).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90099350.
RA   BURD C.G., SWANSON M.S., GOERLACH M., DREYFUSS G.;
RT   "Primary structures of the heterogeneous nuclear ribonucleoprotein
RT   A2, B1, and C2 proteins: a diversity of RNA binding proteins is
RT   generated by small peptide inserts.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 86:9788-9792(1989).
RN   [2]
RP   SEQUENCE OF 1-107 AND 121-303 FROM N.A.
RX   MEDLINE; 87257872.
RA   SWANSON M.S., NAKAGAWA T.Y., LEVAN K., DREYFUSS G.;
RT   "Primary structure of human nuclear ribonucleoprotein particle C
RT   proteins: conservation of sequence and domain structures in
RT   heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins.";
RL   MOL. CELL. BIOL. 7:1731-1739(1987).
RN   [3]
RP   PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX   MEDLINE; 90067819.
RA   MERRILL B.M., BARNETT S.F., LESTOURGEON W.M., WILLIAMS K.R.;
RT   "Primary structure differences between proteins C1 and C2 of HeLa 40S
RT   nuclear ribonucleoprotein particles.";
RL   NUCLEIC ACIDS RES. 17:8441-8449(1989).
```

```
RN   [4]
RP   STRUCTURE BY NMR OF 1-94.
RX   MEDLINE; 92329450.
RA   WITTEKING M., GOERLACH M., FRIEDRICHS M., DREYFUSS G., MUELLER L.;
RT   "1H, 13C, and 15N NMR assignments and global folding pattern of the
RT   RNA-binding domain of the human hnRNP C proteins.";
RL   BIOCHEMISTRY 31:6254-6265(1992).
RN   [5]
RP   STRUCTURE BY NMR OF 1-94.
RX   MEDLINE; 92371436.
RA   GOERLACH M., WITTEKING M., BECKMAN R.A., MUELLER L., DREYFUSS G.;
RT   "Interaction of the RNA-binding domain of the hnRNP C proteins with
RT   RNA.";
RL   EMBO J. 11:3289-3295(1992).
CC   -!- FUNCTION: MAY PLAY A ROLE IN RIBONUCLEOSOME ASSEMBLY BY
CC       NEUTRALIZING BASIC PROTEINS SUCH AS A AND B CORE HNRNP.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; COMPONENT OF RIBONUCLEOSOMES.
CC   -!- PTM: PHOSPHORYLATED (PROBABLE).
CC   -!- ALTERNATIVE PRODUCTS: C1 AND C2 PROTEINS PROBABLY RESULT FROM
CC       ALTERNATIVE SPLICING OF THE PRE-MRNA. C2 IS IDENTICAL TO C1
CC       EXCEPT FOR AN INSERT OF 13 AA (AA 108-120) ONLY PRESENT IN C2.
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
CC   -!- SIMILARITY: HIGH, TO X.LAEVIS PROTEIN C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29063; G337455; -.
DR   EMBL; M16342; G306875; -.
DR   PIR; C34504; C34504.
DR   PIR; A26885; A26885.
DR   AARHUS/GHENT-2DPAGE; 7207; IEF.
DR   AARHUS/GHENT-2DPAGE; 7222; IEF.
DR   MIM; 164020; -.
DR   SWISS-2DPAGE; P07910; HUMAN.
DR   PROSITE; PS00030; RNP_1; 1.
DR   PFAM; PF00076; rrm; 1.
KW   NUCLEAR PROTEIN; RNA-BINDING; RIBONUCLEOPROTEIN; PHOSPHORYLATION;
KW   ALTERNATIVE SPLICING.
FT   DOMAIN       50     57       RNA-BINDING (RNP1).
FT   DOMAIN      155    161       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      181    303       ASP/GLU-RICH (ACIDIC).
FT   VARSPLIC    108    120       MISSING (IN HNRNP C1).
SQ   SEQUENCE   303 AA;  33298 MW;  09615140 CRC32;

  Query Match           67.1%;  Score 55;  DB 1;  Length 303;
  Best Local Similarity  77.8%;  Pred. No. 4.10e+00;
  Matches     7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    118 LSSSFDLDY 126
          ||:|| |||
Qy      4 LSTSFSLDY 12


RESULT    7
ID   SHIA_ECOLI     STANDARD;      PRT;   438 AA.
AC   P76350;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   SHIKIMATE TRANSPORTER.
GN   SHIA.
OS   ESCHERICHIA COLI.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ESCHERICHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 98192527.
RA   WHIPP M.J., CAMAKARIS H., PITTARD A.J.;
RT   "Cloning and analysis of the shiA gene, which encodes the shikimate
RT   transport system of Escherichia coli K-12.";
RL   GENE 209:185-192(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   SCIENCE 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97251358.
RA   ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T.,
RA   ISONO K., KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M.,
RA   MAKINO K., MIKI T., MIZOBUCHI K., MORI H., MORI T., MOTOMURA K.,
RA   NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,
RA   SAITO N., SAMPEI G., SEKI Y., SIVASUNDARAM S., TAGAMI H.,
RA   TAKEDA J., TAKEMOTO K., WADA C., YAMAMOTO Y., HORIUCHI T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA RES. 3:379-392(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U88529; G1850982; -.
DR   EMBL; AE000290; G1788292; -.
DR   EMBL; D90837; G1736645; -.
DR   ECOGENE; EG20205; SHIA.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR   PFAM; PF00083; sugar_tr; 1.
KW   TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE; SYMPORT.
FT   TRANSMEM     28     48       POTENTIAL.
FT   TRANSMEM     64     84       POTENTIAL.
FT   TRANSMEM    109    129       POTENTIAL.
FT   TRANSMEM    133    153       POTENTIAL.
FT   TRANSMEM    168    188       POTENTIAL.
FT   TRANSMEM    193    213       POTENTIAL.
FT   TRANSMEM    255    275       POTENTIAL.
FT   TRANSMEM    287    307       POTENTIAL.
FT   TRANSMEM    333    353       POTENTIAL.
FT   TRANSMEM    387    407       POTENTIAL.
FT   TRANSMEM    411    431       POTENTIAL.
SQ   SEQUENCE   438 AA;  47817 MW;  8773D0DD CRC32;

  Query Match           67.1%;  Score 55;  DB 1;  Length 438;
  Best Local Similarity  50.0%;  Pred. No. 4.10e+00;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    267 YIVTAFALNY 276
          |: |:|:|:|
Qy      3 YLSTSFSLDY 12


RESULT    8
ID   ACSC_MOOTH     STANDARD;      PRT;   445 AA.
AC   Q07340;
DT   01-FEB-1995 (REL. 31, CREATED)
```

```
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CORRINOID/IRON-SULFUR PROTEIN, LARGE SUBUNIT (C/FE-SP).
GN   ACSC.
OS   MOORELLA THERMOACETICA (CLOSTRIDIUM THERMOACETICUM).
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOORELLA GROUP;
OC   MOORELLA.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 93194857.
RA   LU W.-P., SCHIAU I., CUNNINGHAM J.R., RAGSDALE S.W.;
RT   "Sequence and expression of the gene encoding the
RT   corrinoid/iron-sulfur protein from Clostridium thermoaceticum and
RT   reconstitution of the recombinant protein to full activity.";
RL   J. BIOL. CHEM. 268:5605-5614(1993).
RN   [2]
RP   SEQUENCE OF 1-7 FROM N.A., AND SEQUENCE OF 1-18.
RC   STRAIN=DSM 521;
RX   MEDLINE; 89098907.
RA   ROBERTS D.L., JAMES-HAGSTROM J.E., GARVIN D.K., GORST C.M.,
RA   RUNQUIST J.A., BAUR J.R., HAASE F.C., RAGSDALE S.W.;
RT   "Cloning and expression of the gene cluster encoding key proteins
RT   involved in acetyl-CoA synthesis in Clostridium thermoaceticum: CO
RT   dehydrogenase, the corrinoid/Fe-S protein, and methyltransferase.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 86:32-36(1989).
CC   -!- FUNCTION: ACTS AS A METHYL GROUP CARRIER IN THE ANAEROBIC ACETYL-
CC       COA PATHWAY (WOOD/LJUNGDAHL PATHWAY) OF CARBON MONOXIDE AND CARBON
CC       DIOXIDE FIXATION.
CC   -!- SUBUNIT: HETERODIMER OF A LARGE AND SMALL CHAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L07099; G144854; -.
DR   PIR; A46621; A46621.
DR   HSSP; P01088; 1BFA.
KW   COBALT; IRON-SULFUR; 4FE-4S; CARBON DIOXIDE FIXATION.
FT   INIT_MET      0      0
FT   METAL        16     16       IRON-SULFUR (4FE-4S) (PROBABLE).
FT   METAL        19     19       IRON-SULFUR (4FE-4S) (PROBABLE).
FT   METAL        24     24       IRON-SULFUR (4FE-4S) (PROBABLE).
SQ   SEQUENCE   445 AA;  48021 MW;  FA6ACBC6 CRC32;

  Query Match           67.1%;  Score 55;  DB 1;  Length 445;
  Best Local Similarity  60.0%;  Pred. No. 4.10e+00;
  Matches    6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    337 YVTTNFSLTY 346
          |::|:||| |
Qy      3 YLSTSFSLDY 12


RESULT    9
ID   ICE3_CRILO     STANDARD;      PRT;   277 AA.
AC   Q60431;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   APOPAIN PRECURSOR (EC 3.4.22.-) (CYSTEINE PROTEASE CPP32) (YAMA
DE   PROTEIN) (CPP-32) (CASPASE-3) (CASP-3) (SREBP CLEAVAGE ACTIVITY 1)
DE   (SCA-1).
GN   CASP3 OR CPP32.
OS   CRICETULUS LONGICAUDATUS (LONG-TAILED HAMSTER) (CHINESE HAMSTER).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; CRICETINAE; CRICETULUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 96183185.
RA   WANG X., ZELENSKI N.G., YANG J., SAKAI J., BROWN M.S.,
RA   GOLDSTEIN J.L.;
RT   "Cleavage of sterol regulatory element binding proteins (SREBPs) by
RT   CPP32 during apoptosis.";
RL   EMBO J. 15:1012-1020(1996).
CC   -!- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC       RESPONSIBLE FOR APOPTOSIS EXECUTION. AT THE ONSET OF APOPTOSIS IT
CC       PROTEOLYTICALLY CLEAVES POLY(ADP-RIBOSE) POLYMERASE (PARP) AT A
CC       216-ASP-|-GLY-217 BOND. CLEAVES AND ACTIVATES STEROL REGULATORY
CC       ELEMENT BINDING PROTEINS (SREBPS) BETWEEN THE BASIC HELIX-LOOP-
CC       HELIX LEUCINE ZIPPER DOMAIN AND THE MEMBRANE ATTACHMENT DOMAIN.
CC       CLEAVES AND ACTIVATES CASPASE-6, -7 AND -9 (BY SIMILARITY).
CC   -!- SUBUNIT: HETERODIMER OF A 17 KD (P17) AND A 12 KD (P12) SUBUNIT
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- PTM: CLEAVAGE BY GRANZYME B, CASPASE-6, -8 AND -10 GENERATES THE
CC       TWO ACTIVE SUBUNITS. ADDITIONAL PROCESSING OF THE PROPEPTIDES IS
CC       LIKELY DUE TO THE AUTOCATALYTIC ACTIVITY OF THE ACTIVATED
CC       PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
CC       CASPASE-7 PROTEASE AND THE LARGE SUBUNIT OF CPP32 ALSO OCCUR AND
CC       VICE VERSA (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC       CASPASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U27463; G1244444; -.
DR   PROSITE; PS01121; CASPASE_HIS; 1.
DR   PROSITE; PS01122; CASPASE_CYS; 1.
DR   PFAM; PF00655; ICE_p10; 1.
DR   PFAM; PF00656; ICE_p20; 1.
DR   HSSP; P42574; 1PAU.
KW   HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.
FT   PROPEP        1      9       BY SIMILARITY.
FT   PROPEP       10     28       BY SIMILARITY.
FT   CHAIN        29    175       APOPAIN P17 SUBUNIT.
FT   CHAIN       176    277       APOPAIN P12 SUBUNIT.
FT   ACT_SITE    121    121       BY SIMILARITY.
FT   ACT_SITE    163    163       BY SIMILARITY.
SQ   SEQUENCE   277 AA;  31612 MW;  29F1D5A7 CRC32;

  Query Match           65.9%;  Score 54;  DB 1;  Length 277;
  Best Local Similarity  50.0%;  Pred. No. 6.33e+00;
  Matches    5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     32 YLDSSYKMDY 41
          || :|: :||
Qy      3 YLSTSFSLDY 12


RESULT   10
ID   ICE3_RAT       STANDARD;      PRT;   277 AA.
AC   P55213; P70543; Q62993; P97699;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   APOPAIN PRECURSOR (EC 3.4.22.-) (CYSTEINE PROTEASE CPP32) (YAMA
DE   PROTEIN) (CPP-32) (CASPASE-3) (CASP-3) (SREBP CLEAVAGE ACTIVITY 1)
DE   (SCA-1) (LICE) (IRP).
GN   CASP3 OR CPP32.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96358624.
```

```
RA   JUAN T.S.-C., MCNIECE I.K., JENKINS N.A., GILBERT D.J., COPELAND N.G.,
RA   FLETCHER F.A.;
RT   "Molecular characterization of mouse and rat CPP32 beta gene encoding
RT   a cysteine protease resembling interleukin-1 beta converting enzyme
RT   and CED-3.";
RL   ONCOGENE 13:749-755(1996).
RN   [2]
RP   SEQUENCE OF 30-241 FROM N.A.
RC   TISSUE=OVARY;
RX   MEDLINE; 96042508.
RA   FLAWS J.A., KUGU K., TRBOVICH A.M., DESANTI A., TILLY K.I.,
RA   HIRSHFIELD A.N., TILLY J.L.;
RT   "Interleukin-1 beta-converting enzyme-related proteases (IRPs) and
RT   mammalian cell death: dissociation of IRP-induced oligonucleosomal
RT   endonuclease activity from morphological apoptosis in granulosa cells
RT   of the ovarian follicle.";
RL   ENDOCRINOLOGY 136:5042-5053(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97184204.
RA   NI B., WU X., DU Y., SU Y., HAMILTON-BYRD E., ROCKEY P.K.,
RA   ROSTECK P. JR., POIRIER G.G., PAUL S.M.;
RT   "Cloning and expression of a rat brain interleukin-1beta-converting
RT   enzyme (ICE)-related protease (IRP) and its possible role in
RT   apoptosis of cultured cerebellar granule neurons.";
RL   J. NEUROSCI. 17:1561-1569(1997).
RN   [4]
RP   SEQUENCE OF 1-264 FROM N.A.
RA   YAKOVLEV A.G.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC       RESPONSIBLE FOR APOPTOSIS EXECUTION. AT THE ONSET OF APOPTOSIS IT
CC       PROTEOLYTICALLY CLEAVES POLY(ADP-RIBOSE) POLYMERASE (PARP) AT A
CC       216-ASP-|-GLY-217 BOND. CLEAVES AND ACTIVATES STEROL REGULATORY
CC       ELEMENT BINDING PROTEINS (SREBPS) BETWEEN THE BASIC HELIX-LOOP-
CC       HELIX LEUCINE ZIPPER DOMAIN AND THE MEMBRANE ATTACHMENT DOMAIN.
CC       CLEAVES AND ACTIVATES CASPASE-6, -7 AND -9 (BY SIMILARITY).
CC   -!- SUBUNIT: HETERODIMER OF A 17 KD (P17) AND A 12 KD (P12) SUBUNIT
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, LIVER, AND MUSCLE
CC       BUT NOT IN KIDNEY OR TESTIS.
CC   -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN NEURON-ENRICHED REGIONS
CC       OF THE DEVELOPING BRAIN, BUT DOWN-REGULATED TO LOW LEVELS IN THE
CC       ADULT BRAIN.
CC   -!- PTM: CLEAVAGE BY GRANZYME B, CASPASE-6, -8 AND -10 GENERATES THE
CC       TWO ACTIVE SUBUNITS. ADDITIONAL PROCESSING OF THE PROPEPTIDES IS
CC       LIKELY DUE TO THE AUTOCATALYTIC ACTIVITY OF THE ACTIVATED
CC       PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
CC       CASPASE-7 PROTEASE AND THE LARGE SUBUNIT OF CPP32 ALSO OCCUR AND
CC       VICE VERSA (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC       CASPASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U49930; G1518936; -.
DR   EMBL; U34685; G1004371; -.
DR   EMBL; U84410; G1814027; -.
DR   EMBL; U58656; G1381643; -.
DR   PROSITE; PS01121; CASPASE_HIS; 1.
DR   PROSITE; PS01122; CASPASE_CYS; 1.
DR   PFAM; PF00655; ICE_p10; 1.
DR   PFAM; PF00656; ICE_p20; 1.
DR   HSSP; P42574; 1PAU.
KW   HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.
FT   PROPEP        1      9       BY SIMILARITY.
FT   PROPEP       10     28       BY SIMILARITY.
FT   CHAIN        29    175       APOPAIN P17 SUBUNIT.
FT   CHAIN       176    277       APOPAIN P12 SUBUNIT.
FT   ACT_SITE    121    121       BY SIMILARITY.
FT   ACT_SITE    163    163       BY SIMILARITY.
FT   CONFLICT     25     29       KSMDS -> QVD (IN REF. 4).
FT   CONFLICT    170    170       C -> S (IN REF. 2).
FT   CONFLICT    178    178       T -> A (IN REF. 2).
FT   CONFLICT    182    182       M -> V (IN REF. 2).
FT   CONFLICT    187    187       I -> K (IN REF. 2).
FT   CONFLICT    190    190       E -> G (IN REF. 3).
FT   CONFLICT    199    199       T -> S (IN REF. 2).
FT   CONFLICT    211    211       D -> G (IN REF. 2).
FT   CONFLICT    236    236       L -> I (IN REF. 4).
FT   CONFLICT    245    245       T -> M (IN REF. 3).
SQ   SEQUENCE   277 AA;  31491 MW;  0E106140 CRC32;

  Query Match           65.9%;  Score 54;  DB 1;  Length 277;
  Best Local Similarity  50.0%;  Pred. No. 6.33e+00;
  Matches    5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     32 YLDSSYKMDY 41
          || :|: :||
Qy      3 YLSTSFSLDY 12


RESULT   11
ID   ICE3_MOUSE     STANDARD;      PRT;   277 AA.
AC   P70677; O08668;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   APOPAIN PRECURSOR (EC 3.4.22.-) (CYSTEINE PROTEASE CPP32) (YAMA
DE   PROTEIN) (CPP-32) (CASPASE-3) (CASP-3) (SREBP CLEAVAGE ACTIVITY 1)
DE   (SCA-1) (LICE).
GN   CASP3 OR CPP32.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96358624.
RA   JUAN T.S.-C., MCNIECE I.K., JENKINS N.A., GILBERT D.J., COPELAND N.G.,
RA   FLETCHER F.A.;
RT   "Molecular characterization of mouse and rat CPP32 beta gene encoding
RT   a cysteine protease resembling interleukin-1 beta converting enzyme
RT   and CED-3.";
RL   ONCOGENE 13:749-755(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97224429.
RA   MUKASA T., URASE K., MOMOI M.Y., KIMURA I., MOMOI T.;
RT   "Specific expression of CPP32 in sensory neurons of mouse embryos and
RT   activation of CPP32 in the apoptosis induced by a withdrawal of
RT   NGF.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 231:770-774(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/AN;
RX   MEDLINE; 97190206.
RA   VAN DE CRAEN M., VANDENABEELE P., DECLERCQ W., VAN DEN BRANDE I.,
RA   VAN LOO G., MOLEMANS F., SCHOTTE P., VAN CRIEKINGE W., BEYAERT R.,
RA   FIERS W.;
RT   "Characterization of seven murine caspase family members.";
RL   FEBS LETT. 403:61-69(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   FERNANDES-ALNEMRI T., LITWACK G., ALNEMRI E.S.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC       RESPONSIBLE FOR APOPTOSIS EXECUTION. AT THE ONSET OF APOPTOSIS IT
```

```
CC       PROTEOLYTICALLY CLEAVES POLY(ADP-RIBOSE) POLYMERASE (PARP) AT A
CC       216-ASP-|-GLY-217 BOND. CLEAVES AND ACTIVATES STEROL REGULATORY
CC       ELEMENT BINDING PROTEINS (SREBPS) BETWEEN THE BASIC HELIX-LOOP-
CC       HELIX LEUCINE ZIPPER DOMAIN AND THE MEMBRANE ATTACHMENT DOMAIN.
CC       CLEAVES AND ACTIVATES CASPASE-6, -7 AND -9 (BY SIMILARITY).
CC       CLEAVES IL-1 BETA BETWEEN AN ASP AND AN ALA, RELEASING THE MATURE
CC       CYTOKINE WHICH IS INVOLVED IN A VARIETY OF INFLAMMATORY PROCESSES.
CC   -!- SUBUNIT: HETERODIMER OF A 17 KD (P17) AND A 12 KD (P12) SUBUNIT
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- PTM: CLEAVAGE BY GRANZYME B, CASPASE-6, -8 AND -10 GENERATES THE
CC       TWO ACTIVE SUBUNITS. ADDITIONAL PROCESSING OF THE PROPEPTIDES IS
CC       LIKELY DUE TO THE AUTOCATALYTIC ACTIVITY OF THE ACTIVATED
CC       PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
CC       CASPASE-7 PROTEASE AND THE LARGE SUBUNIT OF CPP32 ALSO OCCUR AND
CC       VICE VERSA (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN SPLEEN, LUNG, LIVER,
CC       KIDNEY AND HEART. LOWER EXPRESSION IN BRAIN, SKELETAL MUSCLE AND
CC       TESTIS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC       CASPASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U54803; G1518968; -.
DR   EMBL; U54802; G1518968; JOINED.
DR   EMBL; U49929; G1518934; -.
DR   EMBL; D86352; G1945544; -.
DR   EMBL; D86352; D1022582; -.
DR   EMBL; Y13086; E315505; -.
DR   EMBL; U19522; G2114328; -.
DR   MGD; MGI:107739; CASP3.
DR   PROSITE; PS01121; CASPASE_HIS; 1.
DR   PROSITE; PS01122; CASPASE_CYS; 1.
DR   PFAM; PF00655; ICE_p10; 1.
DR   PFAM; PF00656; ICE_p20; 1.
DR   HSSP; P42574; 1PAU.
KW   HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.
FT   PROPEP        1      9       BY SIMILARITY.
FT   PROPEP       10     28       BY SIMILARITY.
FT   CHAIN        29    175       APOPAIN P17 SUBUNIT.
FT   CHAIN       176    277       APOPAIN P12 SUBUNIT.
FT   ACT_SITE    121    121       BY SIMILARITY.
FT   ACT_SITE    163    163       BY SIMILARITY.
FT   CONFLICT     17     17       E -> G (IN REF. 2).
FT   CONFLICT     51     51       N -> T (IN REF. 2).
FT   CONFLICT     84     84       Q -> E (IN REF. 2).
FT   CONFLICT     95     95       D -> E (IN REF. 2).
FT   CONFLICT     97     97       L -> M (IN REF. 2).
FT   CONFLICT    128    128       Y -> F (IN REF. 2).
FT   CONFLICT    135    135       E -> D (IN REF. 2).
SQ   SEQUENCE   277 AA;  31474 MW;  EA7DCF1F CRC32;

  Query Match           65.9%;  Score 54;  DB 1;  Length 277;
  Best Local Similarity  50.0%;  Pred. No. 6.33e+00;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     32 YLDSSYKMDY 41
          || :|: :||
Qy      3 YLSTSFSLDY 12


RESULT   12
ID   YCR3_YEAST     STANDARD;      PRT;   611 AA.
AC   P25351;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 69.2 KD PROTEIN IN HSP30-PMP1 INTERGENIC REGION.
GN   YCR023C OR YCR23C OR YCR241.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92245758.
RA   BOLLE P.-A., GILLIQUET V., BERBEN G., DUMONT J., HILGER F.;
RT   "The complete sequence of K3B, a 7.9 kb fragment between PGK1 and
RT   CRY1 on chromosome III, reveals the presence of seven open reading
RT   frames.";
RL   YEAST 8:205-213(1992).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59720; E264487; -.
DR   PIR; S19434; S19434.
DR   PIR; S22273; S22273.
KW   HYPOTHETICAL PROTEIN; TRANSPORT; TRANSMEMBRANE.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    153    173       POTENTIAL.
FT   TRANSMEM    200    220       POTENTIAL.
FT   TRANSMEM    354    372       POTENTIAL.
FT   TRANSMEM    414    434       POTENTIAL.
FT   TRANSMEM    443    463       POTENTIAL.
FT   TRANSMEM    543    563       POTENTIAL.
SQ   SEQUENCE   611 AA;  69198 MW;  74CCDBDE CRC32;

  Query Match           65.9%;  Score 54;  DB 1;  Length 611;
  Best Local Similarity  60.0%;  Pred. No. 6.33e+00;
  Matches     6;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     61 SGYLSSSFAL 70
          ::|||:||:|
Qy      1 ASYLSTSFSL 10


RESULT   13
ID   RL6B_YEAST     STANDARD;      PRT;   175 AA.
AC   P05739;
DT   01-NOV-1988 (REL. 09, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   60S RIBOSOMAL PROTEIN L6-B (L17) (YL16) (RP18).
GN   RPL6B OR YL16B OR RPL16B OR YLR448W OR L9324.4.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93003323.
RA   HASHIMOTO T., SUZUKI K., MIZUTA K., OTAKA E.;
RT   "Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the
RT   two genes encoding Saccharomyces cerevisiae YL16.";
RL   BIOCHIM. BIOPHYS. ACTA 1132:195-198(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA   FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J.,
RA   KUCABA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M.,
RA   JOHNSON D., JOHNSTON L., LANGSTON Y., LATREILLE P., LE T.,
```

```
RA   MARDIS E., MENEZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,
RA   RIFKEN L., RILES L., TAICH A., TREVASKIS E., VIGNATI D.,
RA   WILCOX L., WOHLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
RL   SUBMITTED (MAR-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE OF 1-24.
RA   OTAKA E., HIGO K.-I., ITOH T.;
RT   "Yeast ribosomal proteins: VIII. Isolation of two proteins and
RT   sequence characterization of twenty-four proteins from cytoplasmic
RT   ribosomes.";
RL   MOL. GEN. GENET. 195:544-546(1984).
CC   -!- THERE ARE TWO GENES FOR L6 IN YEAST.
CC   -!- SIMILARITY: BELONGS TO THE L6E FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10226; G218511; -.
DR   EMBL; U22382; G717063; -.
DR   PIR; S11257; S11257.
DR   PIR; S28945; S28945.
DR   SGD; L0001714; RPL6B.
DR   PROSITE; PS01170; RIBOSOMAL_L6E; 1.
DR   PFAM; PF01159; L6e; 1.
KW   RIBOSOMAL PROTEIN; MULTIGENE FAMILY.
FT   INIT_MET      0      0
FT   CONFLICT     67     67       P -> L (IN REF. 1).
SQ   SEQUENCE   175 AA;  19855 MW;  EADAB462 CRC32;

  Query Match           64.6%;  Score 53;  DB 1;  Length 175;
  Best Local Similarity  87.5%;  Pred. No. 9.73e+00;
  Matches     7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    157 YLSASFSL 164
          |||:||||
Qy      3 YLSTSFSL 10


RESULT   14
ID   RL6A_YEAST     STANDARD;      PRT;   175 AA.
AC   Q02326;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   60S RIBOSOMAL PROTEIN L6-A (L17) (YL16) (RP18).
GN   RPL6A OR YL16A OR YML073C.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93003323.
RA   HASHIMOTO T., SUZUKI K., MIZUTA K., OTAKA E.;
RT   "Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the
RT   two genes encoding Saccharomyces cerevisiae YL16.";
RL   BIOCHIM. BIOPHYS. ACTA 1132:195-198(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   BROWN D., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.;
RL   SUBMITTED (OCT-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- THERE ARE TWO GENES FOR L6 IN YEAST.
CC   -!- SIMILARITY: BELONGS TO THE L6E FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10225; G218509; -.
DR   EMBL; Z46373; G914877; -.
DR   PIR; S28944; S28944.
DR   SGD; L0004165; RPL6A.
DR   PROSITE; PS01170; RIBOSOMAL_L6E; 1.
DR   PFAM; PF01159; L6e; 1.
KW   RIBOSOMAL PROTEIN; MULTIGENE FAMILY.
FT   INIT_MET      0      0
SQ   SEQUENCE   175 AA;  19830 MW;  2E155789 CRC32;

  Query Match           64.6%;  Score 53;  DB 1;  Length 175;
  Best Local Similarity  87.5%;  Pred. No. 9.73e+00;
  Matches     7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    157 YLSASFSL 164
          |||:||||
Qy      3 YLSTSFSL 10


RESULT   15
ID   GSHB_ANASP     STANDARD;      PRT;   324 AA.
AC   P45480; Q43879;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   GLUTATHIONE SYNTHETASE (EC 6.3.2.3) (GLUTATHIONE SYNTHASE) (GSH
DE   SYNTHETASE) (GSH-S).
GN   GSHB OR GSH-II.
OS   ANABAENA SP. (STRAIN PCC 7120).
OC   BACTERIA; CYANOBACTERIA; NOSTOCALES; NOSTOCACEAE; ANABAENA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96001250.
RA   DOHERTY H.M., ADAMS D.G.;
RT   "Cloning and sequence of ftsz and flanking regions from the
RT   cyanobacterium Anabaena PCC 7120.";
RL   GENE 163:93-96(1995).
RN   [2]
RP   SEQUENCE OF 24-324 FROM N.A.
RX   MEDLINE; 96099685.
RA   ZHANG C.C., HUGUENIN S., FRIRY A.;
RT   "Analysis of genes encoding the cell division protein FtsZ and a
RT   glutathione synthetase homologue in the cyanobacterium Anabaena sp.
RT   PCC 7120.";
RL   RES. MICROBIOL. 146:445-455(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + GAMMA-L-GLUTAMYL-L-CYSTEINE + GLYCINE =
CC       ADP + ORTHOPHOSPHATE + GLUTATHIONE.
CC   -!- PATHWAY: SECOND STEP IN GLUTATHIONE BIOSYNTHESIS.
CC   -!- SIMILARITY: TO OTHER PROKARYOTIC GSH SYNTHASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14408; G555916; -.
DR   EMBL; Z31371; E190151; -.
DR   HSSP; P04425; 2GLT.
KW   GLUTATHIONE BIOSYNTHESIS; LIGASE; ATP-BINDING.
FT   CONFLICT     50     50       C -> W (IN REF. 2).
FT   CONFLICT    243    243       M -> V (IN REF. 2).
SQ   SEQUENCE   324 AA;  35870 MW;  61CC0D40 CRC32;

  Query Match           64.6%;  Score 53;  DB 1;  Length 324;
  Best Local Similarity  45.5%;  Pred. No. 9.73e+00;
  Matches     5;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;
```

```
Db    102 AYLYATYVLDY 112
          :|| ::: |||
Qy      2 SYLSTSFSLDY 12
```

Search completed: Thu Sep 2 12:39:34 1999
Job time : 9 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:38:52 1999; MasPar time 3.15 Seconds
152.742 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-31
Description: (1-12) from US08599226.pep
Perfect Score: 82
Sequence: 1 ASYLSTSFSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 25.698; Variance 36.778; scale 0.699

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	63	76.8	272	2	D64155	hypothetical protein	3.53e-01
2	62	75.6	1294	2	S58307	hypothetical protein	5.45e-01
3	57	69.5	300	2	JC5714	G protein-coupled rec	4.54e+00
4	56	68.3	315	2	S76267	hypothetical protein	6.85e+00
5	56	68.3	575	2	I59327	olfactory cyclic nucl	6.85e+00
6	55	67.1	303	2	C34504	heterogeneous ribonuc	1.03e+01
7	55	67.1	357	2	G70010	aspartate aminotransf	1.03e+01
8	55	67.1	438	2	G64962	shikimate transport p	1.03e+01
9	55	67.1	446	1	A46621	corrinoid/iron-sulfur	1.03e+01
10	54	65.9	84	2	S57545	probable membrane pro	1.53e+01
11	54	65.9	193	2	S67699	probable membrane pro	1.53e+01
12	54	65.9	212	2	I67437	cysteine proteinase (	1.53e+01
13	54	65.9	220	2	B71811	hypothetical protein	1.53e+01
14	54	65.9	262	2	E64708	conserved hypothetica	1.53e+01
15	54	65.9	277	2	JC5410	CPP32 protein - mouse	1.53e+01
16	54	65.9	277	2	S64710	cysteine proteinase (	1.53e+01
17	54	65.9	611	2	S19434	probable transport pr	1.53e+01
18	53	64.6	176	2	S28944	ribosomal protein L6.	2.28e+01
19	53	64.6	176	2	S55970	ribosomal protein L6.	2.28e+01
20	53	64.6	324	2	JC4291	glutathione synthase	2.28e+01
21	53	64.6	738	1	S51380	protein kinase STE11	2.28e+01
22	52	63.4	122	2	S64359	hypothetical protein	3.36e+01
23	52	63.4	227	2	S73905	CDPdiacylglycerol--gl	3.36e+01
24	52	63.4	248	2	B71036	hypothetical protein	3.36e+01
25	52	63.4	324	2	S51001	transforming protein	3.36e+01
26	52	63.4	336	2	H27743	papG adhesin protein	3.36e+01
27	52	63.4	336	2	S25226	papG protein - Escher	3.36e+01
28	52	63.4	336	2	JC1296	papG fimbrial protein	3.36e+01
29	52	63.4	336	2	I27743	papG adhesin protein	3.36e+01
30	52	63.4	450	2	F64530	conserved hypothetica	3.36e+01
31	52	63.4	450	2	B71976	hypothetical protein	3.36e+01
32	52	63.4	458	2	S75328	DNA-directed DNA poly	3.36e+01
33	52	63.4	475	2	S55093	hypothetical protein	3.36e+01
34	52	63.4	588	2	B70618	probable PE protein -	3.36e+01
35	52	63.4	649	2	G71530	probable ATP synthase	3.36e+01
36	52	63.4	1195	2	A59016	DNA-directed DNA poly	3.36e+01
37	51	62.2	53	2	S31032	gene 87 protein - Myc	4.93e+01
38	51	62.2	118	2	T01883	hypothetical protein	4.93e+01
39	51	62.2	233	2	C70106	conserved hypothetica	4.93e+01
40	51	62.2	381	2	A33477	protein-glutamine gam	4.93e+01
41	51	62.2	552	2	JC1173	cycloheximide resista	4.93e+01
42	51	62.2	715	2	B43943	ATP-dependent membran	4.93e+01
43	51	62.2	817	1	TGHUM1	protein-glutamine gam	4.93e+01
44	51	62.2	1978	2	S77257	hypothetical protein	4.93e+01
45	51	62.2	3396	1	A42551	genome polyprotein -	4.93e+01

ALIGNMENTS

RESULT 1
ENTRY D64155 #type complete
TITLE hypothetical protein HI0597 - Haemophilus influenzae (strain Rd KW20)
ORGANISM #formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
ACCESSIONS D64155
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal Science (1995) 269:496-512
#title Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
#cross-references MUID:95350630
#accession D64155
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-272 ##label TIGR
##cross-references GB:U32741; GB:L42023; NID:g1573582; PID:g1573586; TIGR:HI0597
##note best homolog was a hypothetical protein from Escherichia coli
CLASSIFICATION #superfamily hypothetical protein HI0597
SUMMARY #length 272 #molecular-weight 30523 #checksum 8761

Query Match 76.8%; Score 63; DB 2; Length 272;
Best Local Similarity 66.7%; Pred. No. 3.53e-01;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Db    261 ARYLRTQFGLDY 272
          | || | |:|||
Qy      1 ASYLSTSFSLDY 12
```

RESULT 2
ENTRY S58307 #type complete
TITLE hypothetical protein SPAC18B11.11 - fission yeast (Schizosaccharomyces pombe)

```
ORGANISM        #formal_name Schizosaccharomyces pombe
DATE            14-Jan-1996 #sequence_revision 01-Mar-1996 #text_change
                  31-Oct-1997
ACCESSIONS      S58307
REFERENCE       S58300
   #authors     Devlin, K.; Churcher, C.M.
   #submission  submitted to the EMBL Data Library, August 1995
   #accession   S58307
      ##molecule_type DNA
      ##residues      1-1294 ##label DEV
      ##cross-references EMBL:Z50728; NID:g929886; PID:g929897
GENETICS
   #map_position 1L
SUMMARY         #length 1294  #molecular-weight 149191  #checksum 7509

  Query Match            75.6%;  Score 62;  DB 2;  Length 1294;
  Best Local Similarity  72.7%;  Pred. No. 5.45e-01;
  Matches     8;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    456 SYFDTSFSLDF 466
          ||: ||||||:
Qy      2 SYLSTSFSLDY 12


RESULT    3
ENTRY           JC5714     #type complete
TITLE           G protein-coupled receptor 40 - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            03-Dec-1997 #sequence_revision 23-Jan-1998 #text_change
                  17-Mar-1999
ACCESSIONS      JC5714
REFERENCE       JC5714
   #authors     Sawzdargo, M.; George, S.R.; Nguyen, T.; Xu, S.; Kolakowski
                  Jr., L.F.; O'Dowd, B.F.
   #journal     Biochem. Biophys. Res. Commun. (1997) 239:543-547
   #title       A cluster of four novel human G protein-coupled receptor
                  genes occurring in close proximity to CD22 gene on
                  chromosome 19q13.1.
   #cross-references MUID:98008875
   #accession   JC5714
      ##status        nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-300 ##label SAW
      ##cross-references GB:AF024687; NID:g2612945; PID:g2612946
KEYWORDS        glycoprotein; lipoprotein; thiolester bond
FEATURE
   13-32             #domain transmembrane #status predicted #label TM1\
   43-64             #domain transmembrane #status predicted #label TM2\
   83-102            #domain transmembrane #status predicted #label TM3\
   125-144           #domain transmembrane #status predicted #label TM4\
   184-206           #domain transmembrane #status predicted #label TM5\
   222-243           #domain transmembrane #status predicted #label TM6\
   260-280           #domain transmembrane #status predicted #label TM7\
   155,165           #binding_site carbohydrate (Asn) (covalent) #status
                       predicted\
   289               #binding_site palmitate (Cys) (covalent) #status
                       predicted
SUMMARY         #length 300  #molecular-weight 31457  #checksum 7805

  Query Match            69.5%;  Score 57;  DB 2;  Length 300;
  Best Local Similarity  50.0%;  Pred. No. 4.54e+00;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    105 YLGAAFPLGY 114
          ||:::|:| |
Qy      3 YLSTSFSLDY 12


RESULT    4
ENTRY           S76267     #type complete
TITLE           hypothetical protein - Synechocystis sp. (strain PCC 6803)
ORGANISM        #formal_name Synechocystis sp.
   #variety     PCC 6803
DATE            25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                  21-Aug-1998
ACCESSIONS      S76267
REFERENCE       S74322
   #authors     Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                  Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                  Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                  Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                  S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                  Yasuda, M.; Tabata, S.
   #journal     DNA Res. (1996) 3:109-136
   #title       Sequence analysis of the genome of the unicellular
                  cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                  determination of the entire genome and assignment of
                  potential protein-coding regions.
   #cross-references MUID:97061201
   #accession   S76267
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-315 ##label KAN
      ##cross-references EMBL:D64000; GB:AB001339; NID:g1001484; PID:d1010770;
                          PID:g1001494
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, June 1996
SUMMARY         #length 315  #molecular-weight 34667  #checksum 4918

  Query Match            68.3%;  Score 56;  DB 2;  Length 315;
  Best Local Similarity  60.0%;  Pred. No. 6.85e+00;
  Matches     6;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Db    155 YLTIRFQLDY 164
          ||:  | |||
Qy      3 YLSTSFSLDY 12


RESULT    5
ENTRY           I59327     #type complete
TITLE           olfactory cyclic nucleotide gated cation channel - rat
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                  25-Apr-1997
ACCESSIONS      I59327; I58165
REFERENCE       I59327
   #authors     Bradley, J.; Li, J.; Davidson, N.; Lester, H.A.; Zinn, K.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1994) 91:8890-8894
   #title       Heteromeric olfactory cyclic nucleotide-gated channels: A new
                  subunit that confers increased sensitivity to cAMP.
   #cross-references MUID:94377458
   #accession   I59327
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-575 ##label RES
      ##cross-references EMBL:U12623; NID:g538128; PID:g538129
REFERENCE       I58165
   #authors     Liman, E.R.; Buck, L.B.
   #journal     Neuron (1994) 13:611-621
   #title       A second subunit of the olfactory cyclic nucleotide-gated
                  channel confers high sensitivity to cAMP.
   #cross-references MUID:95000663
   #accession   I58165
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-575 ##label RE2
      ##cross-references EMBL:U12425; NID:g548083; PID:g548084
CLASSIFICATION  #superfamily cAMP receptor protein cyclic nucleotide-binding
                  domain homology
FEATURE
   348-472           #domain cAMP receptor protein cyclic nucleotide-binding
                       domain homology #label CA2
SUMMARY         #length 575  #molecular-weight 65674  #checksum 1183

  Query Match            68.3%;  Score 56;  DB 2;  Length 575;
  Best Local Similarity  63.6%;  Pred. No. 6.85e+00;
```

```
  Matches     7;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Db      64 SYLVAWFVLDY 74
           ||| : | |||
Qy       2 SYLSTSFSLDY 12


RESULT    6
ENTRY             C34504     #type complete
TITLE             heterogeneous ribonuclear particle protein C2 - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change
                    10-Sep-1997
ACCESSIONS        C34504
REFERENCE         A34504
   #authors       Burd, C.G.; Swanson, M.S.; Goerlach, M.; Dreyfuss, G.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1989) 86:9788-9792
   #title         Primary structures of the heterogeneous nuclear
                    ribonucleoprotein A2, B1, and C2 proteins: a diversity of
                    RNA binding proteins is generated by small peptide inserts.
   #cross-references MUID:90099350
   #accession    C34504
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-303 ##label BUR
      ##cross-references GB:M29063; NID:g337454; PID:g337455
CLASSIFICATION    #superfamily unassigned ribonucleoprotein repeat-containing
                    proteins; ribonucleoprotein repeat homology
KEYWORDS          alternative splicing
FEATURE
   17-77              #domain ribonucleoprotein repeat homology #label RRM1
SUMMARY           #length 303  #molecular-weight 33298  #checksum 1186

  Query Match           67.1%;  Score 55;  DB 2;  Length 303;
  Best Local Similarity  77.8%;  Pred. No. 1.03e+01;
  Matches     7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db     118 LSSSFDLDY 126
           ||:|| |||
Qy       4 LSTSFSLDY 12


RESULT    7
ENTRY             G70010     #type complete
TITLE             aspartate aminotransferase homolog yugH - Bacillus subtilis
ORGANISM          #formal_name Bacillus subtilis
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    24-Sep-1998
ACCESSIONS        G70010
REFERENCE         A69580
   #authors       Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                    Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                    Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                    A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                    Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                    Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                    Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                    Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                    Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                    M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                    S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                    Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                    C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                    Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                    Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                    Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                    Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                    Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                    Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                    M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                    M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                    V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                    A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                    Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                    Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                    Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                    Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                    B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                    Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                    Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                    Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                    Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                    Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                    K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                    Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive bacterium
                    Bacillus subtilis.
   #cross-references MUID:98044033
   #accession    G70010
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-357 ##label KUN
      ##cross-references GB:Z99120; GB:AL009126; NID:g2635613; PID:e1184218;
                          PID:g2635636
      ##experimental_source strain 168
GENETICS
   #gene          yugH
CLASSIFICATION    #superfamily aspartate transaminase
SUMMARY           #length 357  #molecular-weight 39461  #checksum 3672

  Query Match           67.1%;  Score 55;  DB 2;  Length 357;
  Best Local Similarity  50.0%;  Pred. No. 1.03e+01;
  Matches     6;  Conservative    1;  Mismatches   5;  Indels   0;  Gaps   0;

Db      75 SRYLSNRFDLSY 86
           : |||  | | |
Qy       1 ASYLSTSFSLDY 12


RESULT    8
ENTRY             G64962     #type complete
TITLE             shikimate transport protein shiA - Escherichia coli
ORGANISM          #formal_name Escherichia coli
DATE              12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
                    01-Feb-1999
ACCESSIONS        G64962; S78630
REFERENCE         A64720
   #authors       Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                    Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                    Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                    Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                    Y.
   #journal       Science (1997) 277:1453-1462
   #title         The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession    G64962
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-438 ##label BLAT
      ##cross-references GB:AE000290; GB:U00096; NID:g2367125; PID:g1788292;
                          UWGP:b1981
      ##experimental_source strain K-12, substrain MG1655
REFERENCE         S78630
   #authors       Whipp, M.J.; Camakaris, H.; Pittard, A.J.
   #journal       Gene (1998) 209:185-192
   #title         Cloning and analysis of the shiA gene, which encodes the
                    shikimate transport system of Escherichia coli K-12.
   #accession    S78630
      ##status        nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues      1-438 ##label WHI
      ##cross-references EMBL:U88529
GENETICS
   #gene          shiA
```

```
FUNCTION
   #description  involved in the uptake of shikimate
   #pathway      aromatic amino acid biosynthesis
CLASSIFICATION   #superfamily citrate utilization determinant
KEYWORDS         transmembrane protein; transport protein
FEATURE
   32-50              #domain transmembrane #status predicted #label TM1\
   58-76              #domain transmembrane #status predicted #label TM2\
   96-114             #domain transmembrane #status predicted #label TM3\
   117-135            #domain transmembrane #status predicted #label TM4\
   171-189            #domain transmembrane #status predicted #label TM5\
   198-216            #domain transmembrane #status predicted #label TM6\
   260-278            #domain transmembrane #status predicted #label TM7\
   291-309            #domain transmembrane #status predicted #label TM8\
   319-337            #domain transmembrane #status predicted #label TM9\
   340-358            #domain transmembrane #status predicted #label TM10\
   390-408            #domain transmembrane #status predicted #label TM11\
   414-432            #domain transmembrane #status predicted #label TM12
SUMMARY          #length 438  #molecular-weight 47817  #checksum 9704

  Query Match           67.1%;  Score 55;  DB 2;  Length 438;
  Best Local Similarity  50.0%;  Pred. No. 1.03e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    267 YIVTAFALNY 276
          |: |:|:|:|
Qy      3 YLSTSFSLDY 12


RESULT    9
ENTRY            A46621     #type complete
TITLE            corrinoid/iron-sulfur protein large chain - Clostridium
                   thermaceticum
ALTERNATE_NAMES  C/Fe-SP large chain AcsC
ORGANISM         #formal_name Clostridium thermaceticum
DATE             21-Sep-1993 #sequence_revision 04-Oct-1996 #text_change
                   18-Sep-1998
ACCESSIONS       A46621
REFERENCE        A46621
   #authors      Lu, W.P.; Schiau, I.; Cunningham, J.R.; Ragsdale, S.W.
   #journal      J. Biol. Chem. (1993) 268:5605-5614
   #title        Sequence and expression of the gene encoding the
                   corrinoid/iron-sulfur protein from Clostridium
                   thermoaceticum and reconstitution of the recombinant
                   protein to full activity.
   #cross-references MUID:93194857
   #accession    A46621
      ##molecule_type DNA; protein
      ##residues      1-446 ##label LU1
      ##note          sequence extracted from NCBI backbone (NCBIN:127877,
                       NCBIP:127878)
GENETICS
   #gene         acsC
CLASSIFICATION   #superfamily corrinoid/iron-sulfur protein large chain
KEYWORDS         4Fe-4S; carbon dioxide fixation; electron transfer;
                   heterodimer; iron-sulfur protein; metalloprotein
FEATURE
   2-446              #product corrinoid/iron-sulfur protein large chain
                        #status experimental #label MAT\
   17,20,25,42        #binding_site 4Fe-4S cluster (Cys) (covalent) #status
                        predicted
SUMMARY          #length 446  #molecular-weight 48153  #checksum 3949

  Query Match           67.1%;  Score 55;  DB 1;  Length 446;
  Best Local Similarity  60.0%;  Pred. No. 1.03e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    338 YVTTNFSLTY 347
          |::|:|||| |
Qy      3 YLSTSFSLDY 12


RESULT   10
ENTRY            S57545     #type complete
TITLE            probable membrane protein YPR012w - yeast (Saccharomyces
                   cerevisiae)
ALTERNATE_NAMES  hypothetical protein YP9531.05
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
                   21-Nov-1997
ACCESSIONS       S57545
REFERENCE        S57541
   #authors      Bowman, S.
   #submission   submitted to the EMBL Data Library, June 1995
   #accession    S57545
      ##molecule_type DNA
      ##residues      1-84 ##label BOW
      ##cross-references EMBL:Z49919; NID:g887584; PID:g887589; MIPS:YPR012w
      ##experimental_source strain AB972
GENETICS
   #map_position 16R
KEYWORDS         transmembrane protein
FEATURE
   11-27              #domain transmembrane #status predicted #label TMM
SUMMARY          #length 84  #molecular-weight 9904  #checksum 7108

  Query Match           65.9%;  Score 54;  DB 2;  Length 84;
  Best Local Similarity  60.0%;  Pred. No. 1.53e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db      6 YLNYAFSLAY 15
          ||: :||| |
Qy      3 YLSTSFSLDY 12


RESULT   11
ENTRY            S67699     #type complete
TITLE            probable membrane protein YDL151c - yeast (Saccharomyces
                   cerevisiae)
ALTERNATE_NAMES  hypothetical protein D1554
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change
                   14-Nov-1997
ACCESSIONS       S67699
REFERENCE        S67693
   #authors      Perea, J.; Blugeon, C.; Delaveau, T.; Jacq, C.
   #submission   submitted to the Protein Sequence Database, July 1996
   #accession    S67699
      ##molecule_type DNA
      ##residues      1-193 ##label PER
      ##cross-references EMBL:Z74199; NID:g1431236; PID:e253062; PID:g1431237;
                          MIPS:YDL151c
      ##experimental_source strain S288C
GENETICS
   #map_position 4L
KEYWORDS         transmembrane protein
FEATURE
   114-130            #domain transmembrane #status predicted #label TMM
SUMMARY          #length 193  #molecular-weight 20893  #checksum 4697

  Query Match           65.9%;  Score 54;  DB 2;  Length 193;
  Best Local Similarity  40.0%;  Pred. No. 1.53e+01;
  Matches     4;  Conservative    6;  Mismatches   0;  Indels   0;  Gaps   0;

Db    149 FLATAFGLNF 158
          :|:|:|:|::
Qy      3 YLSTSFSLDY 12


RESULT   12
ENTRY            I67437     #type fragment
TITLE            cysteine proteinase (EC 3.4.22.-) P32 - rat (fragment)
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
                   17-Mar-1999
ACCESSIONS       I67437
```

```
REFERENCE       I53300
  #authors      Flaws, J.A.; Kugu, K.; Trbovich, A.M.; DeSanti, A.; Tilly,
                  K.I.; Hirshfield, A.N.; Tilly, J.L.
  #journal      Endocrinology (1995) 136:5042-5053
  #title        Interleukin-1-beta-converting enzyme-related proteases (IRPs)
                  and mammalian cell death: dissociation of IRP-induced
                  oligonucleosomal endonuclease activity from morphological
                  apoptosis in granulosa cells of the ovarian follicle.
  #cross-references MUID:96042508
  #accession    I67437
    ##status       preliminary; translated from GB/EMBL/DDBJ
    ##molecule_type mRNA
    ##residues     1-212 ##label RES
    ##cross-references EMBL:U34685; NID:g1004370; PID:g1004371
KEYWORDS        cysteine proteinase; hydrolase
SUMMARY         #length 212  #checksum 472

  Query Match           65.9%;  Score 54;  DB 2;  Length 212;
  Best Local Similarity  50.0%;  Pred. No. 1.53e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db      3 YLDSSYKMDY 12
          || :|: :||
Qy      3 YLSTSFSLDY 12


RESULT   13
ENTRY           B71811     #type complete
TITLE           hypothetical protein jhp1402 - Helicobacter pylori (strain
                  J99)
ORGANISM        #formal_name Helicobacter pylori
  #variety      strain J99
DATE            12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                  05-Mar-1999
ACCESSIONS      B71811
REFERENCE       A71800
  #authors      Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                  Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                  B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                  Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                  Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                  G.F.; Trust, T.J.
  #journal      Nature (1999) 397:176-180
  #title        Genomic sequence comparison of two unrelated isolates of the
                  human gastric pathogen Helicobacter pylori.
  #cross-references MUID:99120557
  #accession    B71811
    ##status       preliminary
    ##molecule_type DNA
    ##residues     1-220 ##label ARN
    ##cross-references GB:AE001562; GB:AE001439; NID:g4156017; PID:g4156023
    ##experimental_source strain J99
GENETICS
  #gene         jhp1402
CLASSIFICATION  #superfamily Escherichia coli ygiH protein
SUMMARY         #length 220  #molecular-weight 23804  #checksum 9738

  Query Match           65.9%;  Score 54;  DB 2;  Length 220;
  Best Local Similarity  58.3%;  Pred. No. 1.53e+01;
  Matches     7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db     83 AVFLSKLFGLDY 94
          | :||  |:|||
Qy      1 ASYLSTSFSLDY 12


RESULT   14
ENTRY           E64708     #type complete
TITLE           conserved hypothetical integral membrane protein HP1509 -
                  Helicobacter pylori (strain 26695)
ORGANISM        #formal_name Helicobacter pylori
DATE            09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
                  13-Nov-1998
ACCESSIONS      E64708
REFERENCE       A64520
  #authors      Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
                  Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
                  H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
                  J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
                  Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
                  McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
                  Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
                  Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
                  Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
                  W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
                  C.M.; Venter, J.C.
  #journal      Nature (1997) 388:539-547
  #title        The complete genome sequence of the gastric pathogen
                  Helicobacter pylori.
  #cross-references MUID:97394467
  #accession    E64708
    ##status       preliminary; nucleic acid sequence not shown;
                    translation not shown
    ##molecule_type DNA
    ##residues     1-262 ##label TOM
    ##cross-references GB:AE000649; GB:AE000511; NID:g2314687; PID:g2314690;
                      TIGR:HP1509
CLASSIFICATION  #superfamily Escherichia coli ygiH protein
SUMMARY         #length 262  #molecular-weight 28835  #checksum 768

  Query Match           65.9%;  Score 54;  DB 2;  Length 262;
  Best Local Similarity  58.3%;  Pred. No. 1.53e+01;
  Matches     7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db    125 AVFLSKLFGLDY 136
          | :||  |:|||
Qy      1 ASYLSTSFSLDY 12


RESULT   15
ENTRY           JC5410     #type complete
TITLE           CPP32 protein - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change
                  17-Mar-1999
ACCESSIONS      JC5410
REFERENCE       JC5410
  #authors      Mukasa, T.; Urase, K.; Momoi, M.Y.; Kimura, I.; Momoi, T.
  #journal      Biochem. Biophys. Res. Commun. (1997) 231:770-774
  #title        Specific expression of CPP32 in sensory neurons of mouse
                  embryos and activation of CPP32 in the apoptosis induced by
                  a withdrawal of NGF.
  #cross-references MUID:97224429
  #accession    JC5410
    ##status       nucleic acid sequence not shown
    ##molecule_type mRNA
    ##residues     1-277 ##label MUK
    ##cross-references DDBJ:D86352
    ##experimental_source embryo
COMMENT     This protein is involved in the apoptosis of dorsal root ganglia
              neurons.
SUMMARY         #length 277  #molecular-weight 31392  #checksum 7414

  Query Match           65.9%;  Score 54;  DB 2;  Length 277;
  Best Local Similarity  50.0%;  Pred. No. 1.53e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     32 YLDSSYKMDY 41
          || :|: :||
Qy      3 YLSTSFSLDY 12


Search completed: Thu Sep  2 12:39:07 1999
Job time : 15 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:40:39 1999; MasPar time 1.39 Seconds
87.779 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-31
Description: (1-12) from US08599226.pep
Perfect Score: 82
Sequence: 1 ASYLSTSFSLDY 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 16.747; Variance 61.471; scale 0.272

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	54	65.9	277	2	US-08-890-	Sequence 2, Applicatio	6.26e+01
2	52	63.4	306	3	PCT-US95-0	Sequence 13, Applicati	9.53e+01
3	49	59.8	267	2	US-08-570-	Sequence 2, Applicatio	1.78e+02
4	49	59.8	898	2	US-08-693-	Sequence 36, Applicati	1.78e+02
5	49	59.8	908	2	US-08-588-	Sequence 3, Applicatio	1.78e+02
6	49	59.8	908	2	US-08-693-	Sequence 33, Applicati	1.78e+02
7	49	59.8	960	2	US-08-693-	Sequence 8, Applicatio	1.78e+02
8	49	59.8	960	2	US-08-588-	Sequence 3, Applicatio	1.78e+02
9	49	59.8	960	2	US-08-355-	Sequence 8, Applicatio	1.78e+02
10	48	58.5	37	3	PCT-US96-0	Sequence 5, Applicatio	2.18e+02
11	48	58.5	117	3	PCT-US96-0	Sequence 11, Applicati	2.18e+02
12	48	58.5	329	1	US-08-348-	Sequence 12, Applicati	2.18e+02
13	48	58.5	343	1	US-08-348-	Sequence 10, Applicati	2.18e+02
14	48	58.5	969	1	US-07-671-	Sequence 4, Applicatio	2.18e+02
15	48	58.5	1174	3	PCT-US95-0	Sequence 29, Applicati	2.18e+02
16	48	58.5	1174	2	US-08-639-	Sequence 29, Applicati	2.18e+02
17	48	58.5	1174	1	US-08-349-	Sequence 29, Applicati	2.18e+02
18	48	58.5	1174	1	US-08-239-	Sequence 29, Applicati	2.18e+02
19	48	58.5	1174	2	US-08-598-	Sequence 29, Applicati	2.18e+02
20	48	58.5	1177	4	5169629-2	Patent No. 5169629.	2.18e+02
21	48	58.5	1177	1	US-07-828-	Sequence 8, Applicatio	2.18e+02
22	48	58.5	1177	3	PCT-US92-1	Sequence 8, Applicatio	2.18e+02
23	48	58.5	1177	1	US-07-920-	Sequence 2, Applicatio	2.18e+02
24	48	58.5	1178	4	5254799-5	Patent No. 5254799.	2.18e+02
25	48	58.5	1178	1	US-08-446-	Sequence 5, Applicatio	2.18e+02
26	48	58.5	1178	2	US-08-463-	Sequence 5, Applicatio	2.18e+02
27	48	58.5	1182	2	US-08-598-	Sequence 34, Applicati	2.18e+02
28	48	58.5	1182	1	US-08-349-	Sequence 34, Applicati	2.18e+02
29	48	58.5	1188	3	PCT-US95-0	Sequence 34, Applicati	2.18e+02
30	48	58.5	1188	1	US-08-239-	Sequence 34, Applicati	2.18e+02
31	48	58.5	1188	2	US-08-639-	Sequence 34, Applicati	2.18e+02
32	47	57.3	180	2	US-08-447-	Sequence 8, Applicatio	2.67e+02
33	47	57.3	180	2	US-08-447-	Sequence 7, Applicatio	2.67e+02
34	47	57.3	192	2	US-08-531-	Sequence 29, Applicati	2.67e+02
35	47	57.3	279	3	PCT-US95-0	Sequence 5, Applicatio	2.67e+02
36	47	57.3	281	2	US-08-810-	Sequence 2, Applicatio	2.67e+02
37	47	57.3	282	1	US-08-118-	Sequence 52, Applicati	2.67e+02
38	47	57.3	282	3	PCT-US93-0	Sequence 52, Applicati	2.67e+02
39	47	57.3	325	4	5320941-2	Patent No. 5320941.	2.67e+02
40	47	57.3	485	1	US-08-068-	Sequence 1, Applicatio	2.67e+02
41	47	57.3	485	1	US-08-464-	Sequence 1, Applicatio	2.67e+02
42	47	57.3	505	1	US-08-068-	Sequence 3, Applicatio	2.67e+02
43	47	57.3	505	1	US-08-464-	Sequence 3, Applicatio	2.67e+02
44	47	57.3	700	3	PCT-US95-0	Sequence 2, Applicatio	2.67e+02
45	47	57.3	1487	2	US-08-760-	Sequence 2, Applicatio	2.67e+02

ALIGNMENTS

```
RESULT    1
ID   US-08-890-542A-2     STANDARD;     PRT;   277 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08890542A
XX
CC    Sequence 2, Application US/08890542A
CC    Patent No. 5840509
CC     GENERAL INFORMATION:
CC       APPLICANT:  Ni, Binhui
CC       APPLICANT:  Paul, Steven M
CC       APPLICANT:  Wu, Xin
CC       TITLE OF INVENTION:  PROTEASE AND RELATED NUCLEIC ACID
CC       TITLE OF INVENTION:  COMPOUNDS
CC       NUMBER OF SEQUENCES:  3
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Eli Lilly and Company
CC         STREET:  Lilly Corporate Center
CC         CITY:  Indianapolis
CC         STATE:  Indiana
CC         COUNTRY:  United States of America
CC         ZIP:  46285
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/890,542A
CC         FILING DATE:  09-JUL-1997
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Gaylo, Paul J.
CC         REGISTRATION NUMBER:  36,808
CC         REFERENCE/DOCKET NUMBER:  X-10704
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (317) 276-0756
CC         TELEFAX:  (317) 276-3861
CC      INFORMATION FOR SEQ ID NO:  2:
CC        SEQUENCE CHARACTERISTICS:
CC          LENGTH:  277 amino acids
CC          TYPE:  amino acid
CC          TOPOLOGY:  linear
CC        MOLECULE TYPE:  protein
```

```
SQ   SEQUENCE   277 AA; 31449 MW; 406172 CN;

  Query Match           65.9%;  Score 54;  DB 2;  Length 277;
  Best Local Similarity  50.0%;  Pred. No. 6.26e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     32 YLDSSYKMDY 41
          || :|: :||
Qy      3 YLSTSFSLDY 12


RESULT    2
ID   PCT-US95-06119-13     STANDARD;     PRT;   306 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 13, Application PC/TUS9506119
XX
CC    Sequence 13, Application PC/TUS9506119
CC     GENERAL INFORMATION:
CC       APPLICANT:
CC       TITLE OF INVENTION:  STREPTOCOCCUS PNEUMONIAE CAPSULAR
CC       TITLE OF INVENTION:  POLYSACCHARIDE GENES AND FLANKING REGIONS
CC       NUMBER OF SEQUENCES:  20
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Arnold, White & Durkee
CC         STREET:  P.O. Box 4433
CC         CITY:  Houston
CC         STATE:  TX
CC         COUNTRY:  United States of America
CC         ZIP:  77210
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
CC         SOFTWARE:  PatentIn Release #1.0, Version
CC         SOFTWARE:  #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/06119
CC         FILING DATE:  CONCURRENTLY HEREWITH
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/243,546
CC         FILING DATE:  16-MAY-1994
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Parker, David L.
CC         REGISTRATION NUMBER:  32,165
CC         REFERENCE/DOCKET NUMBER:  AMCY018P--
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (512) 418-3000
CC         TELEFAX:  (713) 789-2679
CC         TELEX:  79-0924
CC     INFORMATION FOR SEQ ID NO:  13:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  306 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   306 AA; 33942 MW; 497482 CN;

  Query Match           63.4%;  Score 52;  DB 3;  Length 306;
  Best Local Similarity  40.0%;  Pred. No. 9.53e+01;
  Matches     4;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     67 YFDSTFELEY 76
          |: ::| |:|
Qy      3 YLSTSFSLDY 12


RESULT    3
ID   US-08-570-929-2     STANDARD;     PRT;   267 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08570929
XX
CC    Sequence 2, Application US/08570929
CC    Patent No. 5773692
CC     GENERAL INFORMATION:
CC       APPLICANT:
CC       TITLE OF INVENTION:  Anti-sense RNA for CAB Transcript to Reduce
CC       TITLE OF INVENTION:  Chlorophyll Content in Plants
CC       NUMBER OF SEQUENCES:  2
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/570,929
CC         FILING DATE:
CC         CLASSIFICATION:  800
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC        LENGTH: 267 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   267 AA; 28387 MW; 353655 CN;

  Query Match           59.8%;  Score 49;  DB 2;  Length 267;
  Best Local Similarity  50.0%;  Pred. No. 1.78e+02;
  Matches     6;  Conservative    3;  Mismatches   3;  Indels   0;  Gaps   0;

Db     67 PSYLTGEFPGDY 78
          :|||:  |: ||
Qy      1 ASYLSTSFSLDY 12


RESULT    4
ID   US-08-693-697-36     STANDARD;     PRT;   898 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 36, Application US/08693697
XX
CC    Sequence 36, Application US/08693697
CC    Patent No. 5869610
CC     GENERAL INFORMATION:
CC       APPLICANT:  Snodgrass, H. R.
CC       APPLICANT:  Cioffi, Joseph
CC       APPLICANT:  Zupancic, Thomas J.
CC       APPLICANT:  Shafer, Alan W.
CC       TITLE OF INVENTION:  Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
CC       TITLE OF INVENTION:  RECEPTOR
CC       NUMBER OF SEQUENCES:  38
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0b
```

```
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/693,697
CC         FILING DATE:  05-AUG-1996
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Poissant, Brian M.
CC         REGISTRATION NUMBER:  28,462
CC         REFERENCE/DOCKET NUMBER:  8907-0037-999
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  650-493-4935
CC         TELEFAX:  650-493-5556
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  36:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  898 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       FRAGMENT TYPE:  internal
SQ   SEQUENCE   898 AA; 102731 MW; 4634755 CN;

  Query Match           59.8%;  Score 49;  DB 2;  Length 898;
  Best Local Similarity  50.0%;  Pred. No. 1.78e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     19 YVITAFNLSY 28
          |: |:|:| |
Qy      3 YLSTSFSLDY 12


RESULT    5
ID   US-08-588-526-3     STANDARD;     PRT;   908 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08588526
XX
CC    Sequence 3, Application US/08588526
CC    Patent No. 5882860
CC     GENERAL INFORMATION:
CC       APPLICANT:  Snodgrass, H.
CC       APPLICANT:  Cioffi, Joseph
CC       APPLICANT:  Zupancic, Thomas
CC       APPLICANT:  Shafer, Alan
CC       TITLE OF INVENTION:  DETECTION OF A LEPTIN RECEPTOR
CC       TITLE OF INVENTION:  VARIANT
CC       TITLE OF INVENTION:  AND METHODS FOR REGULATING OBESITY
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  US
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/588,526
CC         FILING DATE:  18-JAN-1996
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Poissant, Brian M.
CC         REGISTRATION NUMBER:  28,462
CC         REFERENCE/DOCKET NUMBER:  8907-030
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  908 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   908 AA; 103856 MW; 4758599 CN;

  Query Match           59.8%;  Score 49;  DB 2;  Length 908;
  Best Local Similarity  50.0%;  Pred. No. 1.78e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     19 YVITAFNLSY 28
          |: |:|:| |
Qy      3 YLSTSFSLDY 12


RESULT    6
ID   US-08-693-697-33     STANDARD;     PRT;   908 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 33, Application US/08693697
XX
CC    Sequence 33, Application US/08693697
CC    Patent No. 5869610
CC     GENERAL INFORMATION:
CC       APPLICANT:  Snodgrass, H. R.
CC       APPLICANT:  Cioffi, Joseph
CC       APPLICANT:  Zupancic, Thomas J.
CC       APPLICANT:  Shafer, Alan W.
CC       TITLE OF INVENTION:  Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
CC       TITLE OF INVENTION:  RECEPTOR
CC       NUMBER OF SEQUENCES:  38
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0b
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/693,697
CC         FILING DATE:  05-AUG-1996
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Poissant, Brian M.
CC         REGISTRATION NUMBER:  28,462
CC         REFERENCE/DOCKET NUMBER:  8907-0037-999
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  650-493-4935
CC         TELEFAX:  650-493-5556
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  33:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  908 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       FRAGMENT TYPE:  internal
SQ   SEQUENCE   908 AA; 103702 MW; 4753903 CN;

  Query Match           59.8%;  Score 49;  DB 2;  Length 908;
```

```
  Best Local Similarity  50.0%;  Pred. No. 1.78e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     19 YVITAFNLSY 28
          |: |:|:| |
Qy      3 YLSTSFSLDY 12


RESULT    7
ID   US-08-693-697-8     STANDARD;     PRT;   960 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/08693697
XX
CC    Sequence 8, Application US/08693697
CC    Patent No. 5869610
CC     GENERAL INFORMATION:
CC       APPLICANT:  Snodgrass, H. R.
CC       APPLICANT:  Cioffi, Joseph
CC       APPLICANT:  Zupancic, Thomas J.
CC       APPLICANT:  Shafer, Alan W.
CC       TITLE OF INVENTION:  Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
CC       TITLE OF INVENTION:  RECEPTOR
CC       NUMBER OF SEQUENCES:  38
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0b
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/693,697
CC         FILING DATE:  05-AUG-1996
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Poissant, Brian M.
CC         REGISTRATION NUMBER:  28,462
CC         REFERENCE/DOCKET NUMBER:  8907-0037-999
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  650-493-4935
CC         TELEFAX:  650-493-5556
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  960 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   960 AA; 109634 MW; 5296019 CN;

  Query Match           59.8%;  Score 49;  DB 2;  Length 960;
  Best Local Similarity  50.0%;  Pred. No. 1.78e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     19 YVITAFNLSY 28
          |: |:|:| |
Qy      3 YLSTSFSLDY 12


RESULT    8
ID   US-08-588-190-3     STANDARD;     PRT;   960 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08588190
XX
CC    Sequence 3, Application US/08588190
CC    Patent No. 5856098
CC     GENERAL INFORMATION:
CC       APPLICANT:  Snodgrass, H. Ralph
CC       APPLICANT:  Cioffi, Joseph
CC       APPLICANT:  Zupancic, Thomas Joel
CC       APPLICANT:  Shafer, Alan Wayne
CC       TITLE OF INVENTION:  DETECTION OF A LEPTIN RECEPTOR
CC       TITLE OF INVENTION:  VARIANT AND METHODS FOR REGULATING OBESITY
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds LLP
CC         STREET:  1155 Avenue of The Americas
CC         CITY:  New York
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10036-2811
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/588,190
CC         FILING DATE:  18-JAN-1996
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Poissant, Brian M.
CC         REGISTRATION NUMBER:  28,462
CC         REFERENCE/DOCKET NUMBER:  008907-0029-999
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  650-493-4935
CC         TELEFAX:  650-493-5556
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  960 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   960 AA; 109698 MW; 5295408 CN;

  Query Match           59.8%;  Score 49;  DB 2;  Length 960;
  Best Local Similarity  50.0%;  Pred. No. 1.78e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     19 YVITAFNLSY 28
          |: |:|:| |
Qy      3 YLSTSFSLDY 12


RESULT    9
ID   US-08-355-888A-8     STANDARD;     PRT;   960 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/08355888A
XX
CC    Sequence 8, Application US/08355888A
CC    Patent No. 5763211
CC     GENERAL INFORMATION:
CC       APPLICANT:  Snodgrass, H. R.
CC       APPLICANT:  Cioffi, Joseph
```

```
CC       APPLICANT:  Zupancic, Thomas J.
CC       APPLICANT:  Shafer, Alan W.
CC       TITLE OF INVENTION:  Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
CC       TITLE OF INVENTION:  RECEPTOR
CC       NUMBER OF SEQUENCES:  31
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/355,888A
CC         FILING DATE:  14-DEC-1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Poissant, Brian M.
CC         REGISTRATION NUMBER:  28,462
CC         REFERENCE/DOCKET NUMBER:  7225-078
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-9741/8864
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  960 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   960 AA; 109634 MW; 5296019 CN;

  Query Match           59.8%;  Score 49;  DB 2;  Length 960;
  Best Local Similarity  50.0%;  Pred. No. 1.78e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     19 YVITAFNLSY 28
          |: |:|:| |
Qy      3 YLSTSFSLDY 12


RESULT   10
ID   PCT-US96-08730-5     STANDARD;     PRT;   37 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application PC/TUS9608730
XX
CC    Sequence 5, Application PC/TUS9608730
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cassels, Frederick
CC       APPLICANT:  Anderson, Jeffrey
CC       APPLICANT:  Carter, John Mark
CC       TITLE OF INVENTION:  Methods of Raising Antibodies Against E.
CC       TITLE OF INVENTION:  Coli of the Family CSF-CFA./1
CC       NUMBER OF SEQUENCES:  15
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Glenna Hendricks
CC         STREET:  P.O. Box 2509
CC         CITY:  Fairfax
CC         STATE:  VA
CC         COUNTRY:  USA
CC         ZIP:  22031
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US96/08730
CC         FILING DATE:  03-JUN-1996
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Hendricks, Glenna
CC         REGISTRATION NUMBER:  32,535
CC         REFERENCE/DOCKET NUMBER:  PCT/US96/08730
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703) 591-4470
CC         TELEFAX:  (703) 591-4428
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  37 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  internal
SQ   SEQUENCE   37 AA; 3864 MW; 7776 CN;

  Query Match           58.5%;  Score 48;  DB 3;  Length 37;
  Best Local Similarity  50.0%;  Pred. No. 2.18e+02;
  Matches     6;  Conservative    3;  Mismatches   3;  Indels   0;  Gaps   0;

Db     22 GSYLPTAVELTY 33
          :|||:|:  | |
Qy      1 ASYLSTSFSLDY 12


RESULT   11
ID   PCT-US96-08730-11     STANDARD;     PRT;   117 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 11, Application PC/TUS9608730
XX
CC    Sequence 11, Application PC/TUS9608730
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cassels, Frederick
CC       APPLICANT:  Anderson, Jeffrey
CC       APPLICANT:  Carter, John Mark
CC       TITLE OF INVENTION:  Methods of Raising Antibodies Against E.
CC       TITLE OF INVENTION:  Coli of the Family CSF-CFA./1
CC       NUMBER OF SEQUENCES:  15
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Glenna Hendricks
CC         STREET:  P.O. Box 2509
CC         CITY:  Fairfax
CC         STATE:  VA
CC         COUNTRY:  USA
CC         ZIP:  22031
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US96/08730
CC         FILING DATE:  03-JUN-1996
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Hendricks, Glenna
CC         REGISTRATION NUMBER:  32,535
CC         REFERENCE/DOCKET NUMBER:  PCT/US96/08730
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703) 591-4470
```

```
CC         TELEFAX:  (703) 591-4428
CC     INFORMATION FOR SEQ ID NO:  11:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  117 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  internal
SQ   SEQUENCE   117 AA; 12389 MW; 76297 CN;

  Query Match           58.5%;  Score 48;  DB 3;  Length 117;
  Best Local Similarity  50.0%;  Pred. No. 2.18e+02;
  Matches     6;  Conservative    3;  Mismatches   3;  Indels   0;  Gaps   0;

Db     22 GSYLPTAVELTY 33
          :|||:|:  | |
Qy      1 ASYLSTSFSLDY 12


RESULT   12
ID   US-08-348-792-12     STANDARD;     PRT;   329 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 12, Application US/08348792
XX
CC    Sequence 12, Application US/08348792
CC    Patent No. 5576423
CC     GENERAL INFORMATION:
CC       APPLICANT:  Aversa, Gregorio
CC       APPLICANT:  Chang, Chia-Chun J.
CC       APPLICANT:  Cocks, Benjamin G.
CC       APPLICANT:  de Vries, Jan E.
CC       TITLE OF INVENTION:  PURIFIED GENES ENCODING MAMMALIAN CELL
CC       TITLE OF INVENTION:  SURFACE ANTIGENS; PROTEINS AND ANTIBODIES
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  DNAX Research Institute
CC         STREET:  901 California Avenue
CC         CITY:  Palo Alto
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  94304-1104
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/348,792
CC         FILING DATE:  02-DEC-1994
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Ching, Edwin P.
CC         REGISTRATION NUMBER:  34,090
CC         REFERENCE/DOCKET NUMBER:  DX0436
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-852-9196
CC         TELEFAX:  415-496-1200
CC     INFORMATION FOR SEQ ID NO:  12:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  329 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   329 AA; 36590 MW; 607424 CN;

  Query Match           58.5%;  Score 48;  DB 1;  Length 329;
  Best Local Similarity  50.0%;  Pred. No. 2.18e+02;
  Matches     5;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db     14 FLSLAFELSY 23
          :|| :| | |
Qy      3 YLSTSFSLDY 12


RESULT   13
ID   US-08-348-792-10     STANDARD;     PRT;   343 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 10, Application US/08348792
XX
CC    Sequence 10, Application US/08348792
CC    Patent No. 5576423
CC     GENERAL INFORMATION:
CC       APPLICANT:  Aversa, Gregorio
CC       APPLICANT:  Chang, Chia-Chun J.
CC       APPLICANT:  Cocks, Benjamin G.
CC       APPLICANT:  de Vries, Jan E.
CC       TITLE OF INVENTION:  PURIFIED GENES ENCODING MAMMALIAN CELL
CC       TITLE OF INVENTION:  SURFACE ANTIGENS; PROTEINS AND ANTIBODIES
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  DNAX Research Institute
CC         STREET:  901 California Avenue
CC         CITY:  Palo Alto
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  94304-1104
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/348,792
CC         FILING DATE:  02-DEC-1994
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Ching, Edwin P.
CC         REGISTRATION NUMBER:  34,090
CC         REFERENCE/DOCKET NUMBER:  DX0436
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-852-9196
CC         TELEFAX:  415-496-1200
CC     INFORMATION FOR SEQ ID NO:  10:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  343 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   343 AA; 38094 MW; 663822 CN;

  Query Match           58.5%;  Score 48;  DB 1;  Length 343;
  Best Local Similarity  50.0%;  Pred. No. 2.18e+02;
  Matches     5;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db     14 FLSLAFELSY 23
          :|| :| | |
Qy      3 YLSTSFSLDY 12


RESULT   14
ID   US-07-671-817A-4     STANDARD;     PRT;   969 AA.
XX
AC   xxxxxx
XX
DT
```

```
XX
DE   Sequence 4, Application US/07671817A
XX
CC    Sequence 4, Application US/07671817A
CC    Patent No. 5424409
CC     GENERAL INFORMATION:
CC       APPLICANT:  Ely, Susan
CC       APPLICANT:  Tippett, Janet M
CC       TITLE OF INVENTION:  DNA constructs
CC       NUMBER OF SEQUENCES:  6
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cushman, Darby and Cushman
CC         STREET:  Eleventh floor, 1615 L Street, N.W.
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20036-3601
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/671,817A
CC         FILING DATE:  19910401
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  GB 8823068.5
CC         FILING DATE:  30-SEP-1988
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/GB89/01157
CC         FILING DATE:  29-SEP-1989
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202) 861-3000
CC         TELEFAX:  (202) 822-0944
CC         TELEX:  6714627 CUSH
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  969 amino acids
CC         TYPE:  AMINO ACID
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   969 AA; 109598 MW; 4754786 CN;

  Query Match           58.5%;  Score 48;  DB 1;  Length 969;
  Best Local Similarity  60.0%;  Pred. No. 2.18e+02;
  Matches     6;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Db    653 TYLSDEFCLD 662
          :|||  | ||
Qy      2 SYLSTSFSLD 11


RESULT   15
ID   PCT-US95-05431-29     STANDARD;     PRT;   1174 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 29, Application PC/TUS9505431
XX
CC    Sequence 29, Application PC/TUS9505431
CC     GENERAL INFORMATION:
CC       APPLICANT:
CC       APPLICANT:  Street address:     5501 Oberlin Drive
CC       APPLICANT:  City:   San Diego
CC       APPLICANT:  State/Province:     California
CC       APPLICANT:  Country:       US
CC       APPLICANT:  Postal code/Zip:    92121
CC       APPLICANT:  Phone number:  (619) 453-8030      Fax number:(619)453-6991
CC       APPLICANT:  Telex number:
CC       TITLE OF INVENTION:  Improvement of Delta-Endotoxin Expression in
CC       TITLE OF INVENTION:  Pseudomonas fluorescens
CC       NUMBER OF SEQUENCES:  34
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  David R. Saliwanchik
CC         STREET:  2421 N.W. 41st Street, Suite A-1
CC         CITY:  Gainesville
CC         STATE:  Florida
CC         COUNTRY:  USA
CC         ZIP:  32606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/05431
CC         FILING DATE:
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Saliwanchik, David R.
CC         REGISTRATION NUMBER:  31,794
CC         REFERENCE/DOCKET NUMBER:  MA83
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (904) 375-8100
CC         TELEFAX:  (904) 372-5800
CC     INFORMATION FOR SEQ ID NO:  29:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1174 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1174 AA; 133119 MW; 7052059 CN;

  Query Match           58.5%;  Score 48;  DB 3;  Length 1174;
  Best Local Similarity  60.0%;  Pred. No. 2.18e+02;
  Matches     6;  Conservative    1;  Mismatches   3;  Indels   0;  Gaps   0;

Db    646 TYLSDEFCLD 655
          :|||  | ||
Qy      2 SYLSTSFSLD 11


Search completed: Thu Sep  2 12:40:46 1999
Job time : 7 secs.
```

MPSRCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

n on: Thu Sep 2 12:41:55 1999; MasPar time 3.46 Seconds
73.821 Million cell updates/sec

bular output not generated.

Title: >US-08-599-226-32
Description: (1-12) from US08599226.pep
Perfect Score: 78
Sequence: 1 ASYLSTSSSLHY 12

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.195; Variance 56.541; scale 0.304

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	78	100.0	12	27	W27591	Anti-TNF-alpha antibo	1.96e-01
2	69	88.5	12	27	W27588	Anti-TNF-alpha antibo	1.85e+00
3	67	85.9	12	27	W27593	Anti-TNF-alpha antibo	3.03e+00
4	64	82.1	12	27	W27592	Anti-TNF-alpha antibo	6.27e+00
5	63	80.8	121	27	W27569	Anti-TNF-alpha antibo	7.99e+00
6	61	78.2	12	27	W27590	Anti-TNF-alpha antibo	1.29e+01
7	57	73.1	12	27	W27586	Anti-TNF-alpha antibo	3.33e+01
8	57	73.1	12	27	W27589	Anti-TNF-alpha antibo	3.33e+01
9	57	73.1	12	27	W27587	Anti-TNF-alpha antibo	3.33e+01
10	52	66.7	875	28	W34987	Bankia gouldi endoglu	1.06e+02
11	51	65.4	12	27	W27594	Anti-TNF-alpha antibo	1.33e+02
12	51	65.4	12	27	W27563	Anti-TNF-alpha antibo	1.33e+02
13	51	65.4	321	23	W21664	Rat spermatid chemore	1.33e+02
14	51	65.4	321	23	W21665	Rat spermatid chemore	1.33e+02
15	50	64.1	417	29	W41941	A. thaliana truncated	1.67e+02
16	50	64.1	444	30	W47364	Arabidopsis thaliana	1.67e+02
17	50	64.1	445	26	W31506	Arabidopsis p-hydroxy	1.67e+02
18	50	64.1	445	29	W41942	Correct A. thaliana p	1.67e+02
19	49	62.8	159	29	W55260	H. pylori ORF 06cp306	2.09e+02
20	49	62.8	177	29	W55472	H. pylori ORF 06cp306	2.09e+02
21	49	62.8	445	29	W41939	Incorrect A. thaliana	2.09e+02
22	49	62.8	491	14	R79382	Mammalian phospholipi	2.09e+02
23	49	62.8	526	29	W55688	H. pylori ORF 09cp107	2.09e+02
24	49	62.8	1274	39	W89253	Human ALP.	2.09e+02
25	48	61.5	29	14	R74832	CDR3 sequence of TCR	2.62e+02
26	48	61.5	29	25	W33827	CDR3 sequence of TCR	2.62e+02
27	48	61.5	159	37	W80222	Human succinate-ubiqu	2.62e+02
28	48	61.5	166	32	W59345	Human ErbB2 domain 1	2.62e+02
29	48	61.5	191	32	W59354	Human ErbB2 protein f	2.62e+02
30	48	61.5	362	16	R48695	G-protein coupled hum	2.62e+02
31	48	61.5	362	19	W02667	G-protein coupled hum	2.62e+02
32	48	61.5	388	5	R25698	Murine adrenergic bet	2.62e+02
33	48	61.5	400	6	R32501	Beta-adrenergic recep	2.62e+02
34	48	61.5	400	10	R54992	Murine beta-3 adrener	2.62e+02
35	48	61.5	402	2	R06495	Beta 3 adrenergic rec	2.62e+02
36	48	61.5	405	12	R62515	Bovine beta3 adrenerg	2.62e+02
37	48	61.5	405	34	W44933	Canine beta-3 adrener	2.62e+02
38	48	61.5	407	31	W53847	Human adrenaline beta	2.62e+02
39	48	61.5	408	10	R54991	Human beta-3 adrenerg	2.62e+02
40	48	61.5	782	23	W19764	Her2-GM-CSF immunosti	2.62e+02
41	48	61.5	1046	19	W02156	Periplasmic chitodext	2.62e+02
42	48	61.5	1255	19	W01111	HER-2/neu protein.	2.62e+02
43	48	61.5	1433	8	R39568	Sequence of c-erbB-2	2.62e+02
44	48	61.5	2799	39	W81867	Human tumour suppress	2.62e+02
45	47	60.3	624	2	R08222	Extracellular portion	3.27e+02

ALIGNMENTS

RESULT 1
ID W27591 standard; peptide; 12 AA.
AC W27591;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226. not
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          100.0%;  Score 78;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 1.96e-01;
  Matches   12;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssslhy 12
          ||||||||||||
Qy      1 ASYLSTSSSLHY 12


RESULT    2
ID   W27588 standard; peptide; 12 AA.
AC   W27588;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          88.5%;  Score 69;  DB 27;  Length 12;
  Best Local Similarity 91.7%;  Pred. No. 1.85e+00;
  Matches   11;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 asylstsssldy 12
          |||||||||| |
Qy      1 ASYLSTSSSLHY 12


RESULT    3
ID   W27593 standard; peptide; 12 AA.
AC   W27593;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 75; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          85.9%;  Score 67;  DB 27;  Length 12;
  Best Local Similarity 83.3%;  Pred. No. 3.03e+00;
  Matches   10;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 asylstassley 12
          ||||||:||| |
Qy      1 ASYLSTSSSLHY 12


RESULT    4
ID   W27592 standard; peptide; 12 AA.
AC   W27592;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
```

```
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           82.1%;  Score 64;  DB 27;  Length 12;
  Best Local Similarity  83.3%;  Pred. No. 6.27e+00;
  Matches    10;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

      1 asflstsssley 12
        ||:|||||| |
Qy      1 ASYLSTSSSLHY 12


RESULT    5
ID   W27569 standard; Protein; 121 AA.
AC   W27569;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain variable region.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody;
KW   heavy chain; variable region; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
DR   N-PSDB; T88404.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 16; Page 76; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain variable region.
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   121 AA;

  Query Match           80.8%;  Score 63;  DB 27;  Length 121;
  Best Local Similarity  81.8%;  Pred. No. 7.99e+00;
  Matches     9;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    100 sylstassldy 110
          |||||:||| |
Qy      2 SYLSTSSSLHY 12


RESULT    6
ID   W27590 standard; peptide; 12 AA.
AC   W27590;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 74; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
```

CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 78.2%; Score 61; DB 27; Length 12;
Best Local Similarity 83.3%; Pred. No. 1.29e+01;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 asylstsfsldy 12
||||||| || |
Qy 1 ASYLSTSSSLHY 12

RESULT 7
ID W27586 standard; peptide; 12 AA.
AC W27586;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 72; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 73.1%; Score 57; DB 27; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.33e+01;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 asylstsssl 10
||||||||||
Qy 1 ASYLSTSSSL 10

RESULT 8
ID W27589 standard; peptide; 12 AA.
AC W27589;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 73.1%; Score 57; DB 27; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.33e+01;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 asylstsssl 10
||||||||||
Qy 1 ASYLSTSSSL 10

```
RESULT    9
ID   W27587 standard; peptide; 12 AA.
AC   W27587;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           73.1%;  Score 57;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 3.33e+01;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssl 10
          ||||||||||
Qy      1 ASYLSTSSSL 10


RESULT   10
ID   W34987 standard; Protein; 875 AA.
AC   W34987;
DT   21-MAY-1998  (first entry)
DE   Bankia gouldi endoglucanase.
KW   Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW   biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW   thermostable enzyme; thermophilic; glycosidase.
OS   Bankia gouldi mix (Clone 37GP2).
PN   WO9744361-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08793.
PR   22-MAY-1996; US-651572.
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
PI   Lam DE, Mathur EJ;
DR   WPI; 98-018435/02.
DR   N-PSDB; T94195.
PT   Endoglucanase(s), preferably form archael bacterium, AEPII 1a -
PT   useful to degrade carboxymethylcellulose and hydrolyse of
PT   beta-1,4-glycosidic bonds in cellulose
PS   Claim 1; Fig 1C; 164pp; English.
CC   This protein comprises an endoglucanase of Bankia gouldi mix (clone
CC   37GP2) that is capable of degrading carboxymethylcellulose and of
CC   hydrolysing the beta-1,4-glycosidic bonds in cellulose.  It has
CC   homology to an endoglucanase of archaebacterium AEPII1a (see
CC   W34985).  It can be produced from native cells or from recombinant
CC   host cells, especially prokaryotic host cells transformed with a
CC   plasmid or virus-derived vector including the endoglucanase DNA
CC   (see T94195).  24 Endoglucanases (see W34986-W35008) are claimed.
CC   They can be used to degrade cellulose for the conversion of plant
CC   biomass into fuels and chemicals, for use in detergents, textiles,
CC   animal feed, waste treatment, and in the fruit juice and brewing
CC   industries for the clarification and extraction of juices.
SQ   Sequence   875 AA;

  Query Match           66.7%;  Score 52;  DB 28;  Length 875;
  Best Local Similarity 58.3%;  Pred. No. 1.06e+02;
  Matches     7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db    618 psyltvdsslty 629
          :|||:  ||| |
Qy      1 ASYLSTSSSLHY 12


RESULT   11
ID   W27594 standard; peptide; 12 AA.
AC   W27594;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Disclosure; Page 75; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
```

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           65.4%;  Score 51;  DB 27;  Length 12;
  Best Local Similarity  88.9%;  Pred. No. 1.33e+02;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 sylstassl 10
          |||||:|||
Qy      2 SYLSTSSSL 10


RESULT   12
ID   W27563 standard; peptide; 12 AA.
AC   W27563;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 12
FT                   /label= Tyr, Asn
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 9; Page 65; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           65.4%;  Score 51;  DB 27;  Length 12;
  Best Local Similarity  88.9%;  Pred. No. 1.33e+02;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 sylstassl 10
          |||||:|||
Qy      2 SYLSTSSSL 10


RESULT   13
ID   W21664 standard; Protein; 321 AA.
AC   W21664;
DT   29-SEP-1997  (first entry)
DE   Rat spermatid chemoreceptor D-8.
KW   Sperm receptor; spermatid chemoreceptor; contraceptive; vaccine;
KW   infertility.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   domain          33..57
FT                   /label= TMD1
FT                   /note= "transmembrane domain 1"
FT   domain          64..84
FT                   /label= TMD2
FT                   /note= "transmembrane domain 2"
FT   domain          106..127
FT                   /label= TMD3
FT                   /note= "transmembrane domain 3"
FT   domain          146..164
FT                   /label= TMD4
FT                   /note= "transmembrane domain 4"
FT   domain          203..225
FT                   /label= TMD5
FT                   /note= "transmembrane domain 5"
FT   domain          242..266
FT                   /label= TMD6
FT                   /note= "transmembrane domain 6"
FT   domain          279..298
FT                   /label= TMD7
FT                   /note= "transmembrane domain 7"
PN   WO9717444-A2.
PD   15-MAY-1997.
PF   08-NOV-1996; U18002.
PR   09-NOV-1995; US-556186.
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PI   Ronnett GV,  Ruat M,  Snyder SH,  Walensky L;
DR   WPI; 97-281033/25.
DR   N-PSDB; T72289.
PT   Nucleic acid encoding rat sperm receptors with homology to odorant
PT   receptors - useful as immuno-contraceptives and for diagnosis of
PT   autoimmune infertility
PS   Claim 8; Page 63-64; 96pp; English.
CC   Novel rat spermatid chemoreceptors D-2, D-7, D-8, D-9 and G-X
CC   (W21662-66) represent a new family of putative chemosensory
CC   receptors that show homology to, but are different from, receptors
CC   of the odorant receptor family.  Their amino acid sequences were
CC   deduced from isolated DNA fragments (T72287-91) isolated from rat
CC   genomic DNA.  Recombinant receptor polypeptides can be produced in
CC   host cells for use as immunocontraceptives e.g. in vaccines, to
CC   detect specific antibodies which are indicative of autoimmune
CC   infertility, or to screen for agents that can modulate the
CC   fertility of an animal by stimulating or inhibiting the binding of
CC   ligands to the receptors.  Host cells can also be used as vaccines.
SQ   Sequence   321 AA;

  Query Match           65.4%;  Score 51;  DB 23;  Length 321;
  Best Local Similarity  50.0%;  Pred. No. 1.33e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    129 ymaicsplhy 138
          |::  |:|||
Qy      3 YLSTSSSLHY 12
```

```
RESULT   14
ID   W21665 standard; Protein; 321 AA.
AC   W21665;
DT   29-SEP-1997  (first entry)
DE   Rat spermatid chemoreceptor D-9.
KW   Sperm receptor; spermatid chemoreceptor; contraceptive; vaccine;
KW   infertility.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   domain          33..57
FT                   /label= TMD1
FT                   /note= "transmembrane domain 1"
FT   domain          64..84
FT                   /label= TMD2
FT                   /note= "transmembrane domain 2"
FT   domain          106..127
FT                   /label= TMD3
FT                   /note= "transmembrane domain 3"
FT   domain          146..164
FT                   /label= TMD4
FT                   /note= "transmembrane domain 4"
FT   domain          203..225
FT                   /label= TMD5
FT                   /note= "transmembrane domain 5"
FT   domain          242..266
FT                   /label= TMD6
FT                   /note= "transmembrane domain 6"
FT   domain          279..298
FT                   /label= TMD7
FT                   /note= "transmembrane domain 7"
PN   WO9717444-A2.
PD   15-MAY-1997.
PF   08-NOV-1996; U18002.
PR   09-NOV-1995; US-556186.
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PI   Ronnett GV,  Ruat M,  Snyder SH,  Walensky L;
DR   WPI; 97-281033/25.
DR   N-PSDB; T72290.
PT   Nucleic acid encoding rat sperm receptors with homology to odorant
PT   receptors - useful as immuno-contraceptives and for diagnosis of
PT   autoimmune infertility
PS   Claim 8; Page 64-65; 96pp; English.
CC   Novel rat spermatid chemoreceptors D-2, D-7, D-8, D-9 and G-X
CC   (W21662-66) represent a new family of putative chemosensory
CC   receptors that show homology to, but are different from, receptors
CC   of the odorant receptor family.  Their amino acid sequences were
CC   deduced from isolated DNA fragments (T72287-91) isolated from rat
CC   genomic DNA.  Recombinant receptor polypeptides can be produced in
CC   host cells for use as immunocontraceptives e.g. in vaccines, to
CC   detect specific antibodies which are indicative of autoimmune
CC   infertility, or to screen for agents that can modulate the
CC   fertility of an animal by stimulating or inhibiting the binding of
CC   ligands to the receptors.  Host cells can also be used as vaccines.
SQ   Sequence   321 AA;

  Query Match           65.4%;  Score 51;  DB 23;  Length 321;
  Best Local Similarity  50.0%;  Pred. No. 1.33e+02;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    129 ymaicsplhy 138
          |::  |:|||
Qy      3 YLSTSSSLHY 12


RESULT   15
ID   W41941 standard; Protein; 417 AA.
AC   W41941;
DT   11-JUN-1998  (first entry)
DE   A. thaliana truncated p-hydroxyphenylpyruvate dioxygenase.
KW   p-hydroxyphenylpyruvate dioxygenase; inhibitor tolerance;
KW   overexpression; herbicide.
OS   Arabidopsis thaliana.
PN   WO9749816-A1.
PD   31-DEC-1997.
PF   26-JUN-1997; U11295.
PR   27-JUN-1996; US-021364.
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PI   Gutteridge S, Maxwell CA, Scolnik PA, Wittenbach VA;
DR   WPI; 98-077179/07.
DR   N-PSDB; V04491.
PT   Plant p-hydroxy:phenyl:pyruvate dioxygenase enzyme - used to isolate
PT   compounds that inhibit the p-hydroxyphenylpyruvate dioxygenase rate
PT   of reaction for use as herbicides
PS   Claim 4; Pages 49-51; 72pp; English.
CC   The present sequence is Arabidopsis thaliana truncated
CC   p-hydroxyphenylpyruvate dioxygenase (p-HD). The p-HD cDNA can be
CC   used to impart tolerance to a compound that inhibits the rate of
CC   reaction of p-HD, or to overexpress p-HD. A compound that inhibits
CC   the activity of p-HD can be used as a herbicide.
SQ   Sequence   417 AA;

  Query Match           64.1%;  Score 50;  DB 29;  Length 417;
  Best Local Similarity  58.3%;  Pred. No. 1.67e+02;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     58 asylltsgdlrf 69
          |||| ||: |::
Qy      1 ASYLSTSSSLHY 12


Search completed: Thu Sep  2 12:42:15 1999
Job time : 20 secs.
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

[illegible] on: Thu Sep 2 12:44:21 1999; MasPar time 1.37 Seconds
88.617 Million cell updates/sec
[illegible]abular output not generated.

Title: >US-08-599-226-32
Description: (1-12) from US08599226.pep
Perfect Score: 78
Sequence: 1 ASYLSTSSSLHY 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
~~1:5A_COMB~~ 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 16.037; Variance 54.698; scale 0.293

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	49	62.8	400	2	US-08-103-	Sequence 9, Applicatio	9.84e+01
2	49	62.8	491	1	US-08-462-	Sequence 2, Applicatio	9.84e+01
3	49	62.8	491	1	US-08-176-	Sequence 2, Applicatio	9.84e+01
4	48	61.5	29	1	US-08-066-	Sequence 59, Applicati	1.23e+02
5	48	61.5	158	2	US-08-828-	Sequence 3, Applicatio	1.23e+02
6	48	61.5	159	2	US-08-828-	Sequence 1, Applicatio	1.23e+02
7	48	61.5	362	3	PCT-US93-0	Sequence 16, Applicati	1.23e+02
8	48	61.5	362	1	US-08-118-	Sequence 16, Applicati	1.23e+02
9	48	61.5	388	1	US-08-087-	Sequence 2, Applicatio	1.23e+02
10	48	61.5	400	1	US-07-916-	Sequence 6, Applicatio	1.23e+02
11	48	61.5	400	1	US-08-351-	Sequence 4, Applicatio	1.23e+02
12	48	61.5	400	1	US-08-351-	Sequence 5, Applicatio	1.23e+02
13	48	61.5	400	1	US-07-783-	Sequence 1, Applicatio	1.23e+02
14	48	61.5	402	1	US-08-444-	Sequence 6, Applicatio	1.23e+02
15	48	61.5	402	1	US-08-087-	Sequence 15, Applicati	1.23e+02
16	48	61.5	405	1	US-08-351-	Sequence 2, Applicatio	1.23e+02
17	48	61.5	408	1	US-07-916-	Sequence 2, Applicatio	1.23e+02
18	48	61.5	408	1	US-08-351-	Sequence 3, Applicatio	1.23e+02
19	48	61.5	1046	2	US-08-386-	Sequence 2, Applicatio	1.23e+02
20	48	61.5	1255	2	US-08-414-	Sequence 68, Applicati	1.23e+02
21	48	61.5	1255	2	US-08-625-	Sequence 2, Applicatio	1.23e+02
22	48	61.5	1255	1	US-08-467-	Sequence 68, Applicati	1.23e+02
23	48	61.5	1255	2	US-08-468-	Sequence 68, Applicati	1.23e+02
24	48	61.5	1255	2	US-08-486-	Sequence 68, Applicati	1.23e+02
25	48	61.5	1255	2	US-08-484-	Sequence 8, Applicatio	1.23e+02
26	48	61.5	1255	2	US-08-356-	Sequence 2, Applicatio	1.23e+02
27	47	60.3	37	3	PCT-US96-0	Sequence 5, Applicatio	1.53e+02
28	47	60.3	117	3	PCT-US96-0	Sequence 11, Applicati	1.53e+02
29	46	59.0	142	3	PCT-US94-1	Sequence 1, Applicatio	1.91e+02
30	46	59.0	142	3	PCT-US94-1	Sequence 3, Applicatio	1.91e+02
31	46	59.0	142	3	PCT-US94-1	Sequence 2, Applicatio	1.91e+02
32	46	59.0	142	3	PCT-US94-1	Sequence 4, Applicatio	1.91e+02
33	46	59.0	142	2	US-08-439-	Sequence 5, Applicatio	1.91e+02
34	46	59.0	142	1	US-08-439-	Sequence 3, Applicatio	1.91e+02
35	46	59.0	142	1	US-08-439-	Sequence 5, Applicatio	1.91e+02
36	46	59.0	142	1	US-08-439-	Sequence 1, Applicatio	1.91e+02
37	46	59.0	142	2	US-08-439-	Sequence 3, Applicatio	1.91e+02
38	46	59.0	460	2	US-08-714-	Sequence 7, Applicatio	1.91e+02
39	46	59.0	460	2	US-08-630-	Sequence 7, Applicatio	1.91e+02
40	46	59.0	505	2	US-08-714-	Sequence 4, Applicatio	1.91e+02
41	46	59.0	505	2	US-08-936-	Sequence 2, Applicatio	1.91e+02
42	46	59.0	506	2	US-08-936-	Sequence 8, Applicatio	1.91e+02
43	46	59.0	506	2	US-08-936-	Sequence 8, Applicatio	1.91e+02
44	46	59.0	506	2	US-08-922-	Sequence 8, Applicatio	1.91e+02
45	46	59.0	816	2	US-08-785-	Sequence 8, Applicatio	1.91e+02

ALIGNMENTS

```
RESULT    1
ID   US-08-103-170-9     STANDARD;     PRT;   400 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 9, Application US/08103170
XX
CC    Sequence 9, Application US/08103170
CC    Patent No. 5885824
CC     GENERAL INFORMATION:
CC       APPLICANT:  Yamada, Tadataka
CC       APPLICANT:  Gantz, Ira
CC       TITLE OF INVENTION:  Recombinant Genomic Clones Encoding
CC       TITLE OF INVENTION:  Histamine H1, H2, and H3 Receptors, Methods For Product
CC       TITLE OF INVENTION:  Thereof, and Proteins Encoded Therefrom
CC       NUMBER OF SEQUENCES:  41
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
CC         ADDRESSEE:  P.C.
CC         STREET:  1755 Jefferson Davis Highway, Fourth Floor
CC         CITY:  Arlington
CC         STATE:  Virginia
CC         COUNTRY:  U.S.A.
CC         ZIP:  22202
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/103,170
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/633,060
CC         FILING DATE:  24-DEC-1990
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Lavalleye, Jean-Paul
CC         REGISTRATION NUMBER:  31,451
CC         REFERENCE/DOCKET NUMBER:  2363-017-55
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703)521-4500
CC         TELEFAX:  (703)486-2347
CC         TELEX:  24885 OPAT UR
CC     INFORMATION FOR SEQ ID NO:  9:
```

```
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  400 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Homo sapiens
SQ   SEQUENCE   400 AA; 42804 MW; 785909 CN;

  Query Match           62.8%;  Score 49;  DB 2;  Length 400;
  Best Local Similarity  40.0%;  Pred. No. 9.84e+01;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    134 YLAVTNTLRY 143
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT    2
ID   US-08-462-733-2     STANDARD;     PRT;   491 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08462733
XX
CC    Sequence 2, Application US/08462733
CC    Patent No. 5610019
CC     GENERAL INFORMATION:
CC       APPLICANT:  Day, Joseph R.
CC       APPLICANT:  Albers, John J.
CC       APPLICANT:  Lofton-Day, Catherine E.
CC       APPLICANT:  Adolphson, Janet L.
CC       TITLE OF INVENTION:  Phospholipid Transfer Proteins
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  ZymoGenetics, Inc.
CC         STREET:  4225 Roosevelt Way, N.E.
CC         CITY:  Seattle
CC         STATE:  WA
CC         COUNTRY:  USA
CC         ZIP:  98105
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/462,733
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Parker, Gary E
CC         REGISTRATION NUMBER:  31-648
CC         REFERENCE/DOCKET NUMBER:  93-11D1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  206-547-8080 ext 322
CC         TELEFAX:  206-548-2329
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  491 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   491 AA; 54569 MW; 1262097 CN;

  Query Match           62.8%;  Score 49;  DB 1;  Length 491;
  Best Local Similarity  33.3%;  Pred. No. 9.84e+01;
  Matches    4;  Conservative    6;  Mismatches   2;  Indels   0;  Gaps   0;

Db    453 AGFLTIGADLHF 464
          |::|: :: ||:
Qy      1 ASYLSTSSSLHY 12


RESULT    3
ID   US-08-176-402-2     STANDARD;     PRT;   491 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08176402
XX
CC    Sequence 2, Application US/08176402
CC    Patent No. 5622843
CC     GENERAL INFORMATION:
CC       APPLICANT:  Day, Joseph R.
CC       APPLICANT:  Albers, John J.
CC       APPLICANT:  Lofton-Day, Catherine E.
CC       APPLICANT:  Adolphson, Janet L.
CC       TITLE OF INVENTION:  Phospholipid Transfer Proteins
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  ZymoGenetics, Inc.
CC         STREET:  4225 Roosevelt Way, N.E.
CC         CITY:  Seattle
CC         STATE:  WA
CC         COUNTRY:  USA
CC         ZIP:  98105
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/176,402
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Parker, Gary E
CC         REGISTRATION NUMBER:  31-684
CC         REFERENCE/DOCKET NUMBER:  93-11
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  206-547-8080 ext 322
CC         TELEFAX:  206-548-2329
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  491 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   491 AA; 54569 MW; 1262097 CN;

  Query Match           62.8%;  Score 49;  DB 1;  Length 491;
  Best Local Similarity  33.3%;  Pred. No. 9.84e+01;
  Matches    4;  Conservative    6;  Mismatches   2;  Indels   0;  Gaps   0;

Db    453 AGFLTIGADLHF 464
          |::|: :: ||:
Qy      1 ASYLSTSSSLHY 12


RESULT    4
ID   US-08-066-325-59     STANDARD;     PRT;   29 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 59, Application US/08066325
XX
CC    Sequence 59, Application US/08066325
CC    Patent No. 5667967
CC     GENERAL INFORMATION:
```

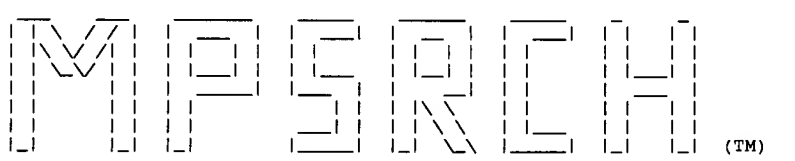
```
CC       APPLICANT:  Steinman, Lawrence
CC       APPLICANT:  Oksenberg, Jorge
CC       APPLICANT:  Bernard, Claude
CC       TITLE OF INVENTION:  T-CELL RECEPTOR VARIABLE TRANSCRIPTS AS DISEASE RELATED MA
CC       NUMBER OF SEQUENCES:  157
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  SEED and BERRY LLP
CC         STREET:  6300 Columbia Center, 701 Fifth Avenue
CC         CITY:  Seattle
CC         STATE:  Washington
CC         COUNTRY:  USA
CC         ZIP:  98104-7092
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/066,325
CC         FILING DATE:  21-MAY-1993
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5667967tenburg Ph.D., Carol
CC         REGISTRATION NUMBER:  39,317
CC         REFERENCE/DOCKET NUMBER:  690068.408C1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (206) 622-4900
CC         TELEFAX:  (206) 682-6031
CC     INFORMATION FOR SEQ ID NO:  59:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  29 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   29 AA; 3155 MW; 4596 CN;

  Query Match           61.5%;  Score 48;  DB 1;  Length 29;
  Best Local Similarity  41.7%;  Pred. No. 1.23e+02;
  Matches     5;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db      4 SSFLGYNSPLHF 15
          :|:|: :|:||:
Qy      1 ASYLSTSSSLHY 12


RESULT    5
ID   US-08-828-832-3     STANDARD;     PRT;   158 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08828832
XX
CC    Sequence 3, Application US/08828832
CC    Patent No. 5827711
CC     GENERAL INFORMATION:
CC       APPLICANT:  Lal, Preeti
CC       APPLICANT:  Shah, Purvi
CC       TITLE OF INVENTION:  NOVEL SUCCINATE DEHYDROGENASE SUBUNIT
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/828,832
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0250 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  158 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  1575011
SQ   SEQUENCE   158 AA; 17096 MW; 135076 CN;

  Query Match           61.5%;  Score 48;  DB 2;  Length 158;
  Best Local Similarity  41.7%;  Pred. No. 1.23e+02;
  Matches     5;  Conservative    4;  Mismatches   3;  Indels   0;  Gaps   0;

Db     81 AAYLNPCSAMDY 92
          |:||:  |:: |
Qy      1 ASYLSTSSSLHY 12


RESULT    6
ID   US-08-828-832-1     STANDARD;     PRT;   159 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08828832
XX
CC    Sequence 1, Application US/08828832
CC    Patent No. 5827711
CC     GENERAL INFORMATION:
CC       APPLICANT:  Lal, Preeti
CC       APPLICANT:  Shah, Purvi
CC       TITLE OF INVENTION:  NOVEL SUCCINATE DEHYDROGENASE SUBUNIT
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/828,832
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
```

```
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0250 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  159 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  Consensus
CC         CLONE:  2454416
SQ   SEQUENCE   159 AA; 17043 MW; 129477 CN;

  Query Match           61.5%;  Score 48;  DB 2;  Length 159;
  Best Local Similarity  41.7%;  Pred. No. 1.23e+02;
  Matches    5;  Conservative    4;  Mismatches   3;  Indels   0;  Gaps   0;

Db     82 AAYLNPCSAMDY 93
          |:||:  |:: |
Qy      1 ASYLSTSSSLHY 12


RESULT    7
ID   PCT-US93-08528-16     STANDARD;     PRT;   362 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 16, Application PC/TUS9308528
XX
CC    Sequence 16, Application PC/TUS9308528
CC     GENERAL INFORMATION:
CC       APPLICANT:  New York University
CC       TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC       TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC       NUMBER OF SEQUENCES:  348
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  BROWDY AND NEIMARK
CC         STREET:  419 Seventh Street, N.W., Suite 300
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/08528
CC         FILING DATE:  09-SEP-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/943,236
CC         FILING DATE:  10-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Townsend, Kevin G.
CC         REGISTRATION NUMBER:  34,033
CC         REFERENCE/DOCKET NUMBER:  MURPHY▫2 PCT
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC         TELEX:  248633
CC     INFORMATION FOR SEQ ID NO:  16:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  362 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   362 AA; 38722 MW; 635563 CN;

  Query Match           61.5%;  Score 48;  DB 3;  Length 362;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     98 YLAVTNPLRY 107
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT    8
ID   US-08-118-270-16     STANDARD;     PRT;   362 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 16, Application US/08118270
XX
CC    Sequence 16, Application US/08118270
CC    Patent No. 5508384
CC     GENERAL INFORMATION:
CC       APPLICANT:  Murphy, Randall B.
CC       APPLICANT:  Schuster, David I.
CC       TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC       TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC       NUMBER OF SEQUENCES:  348
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  BROWDY AND NEIMARK
CC         STREET:  419 Seventh Street, N.W., Suite 300
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/118,270
CC         FILING DATE:  09-SEP-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/943,236
CC         FILING DATE:  10-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Townsend, Kevin G.
CC         REGISTRATION NUMBER:  34,033
CC         REFERENCE/DOCKET NUMBER:  MURPHY▫2A
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC         TELEX:  248633
CC     INFORMATION FOR SEQ ID NO:  16:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  362 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   362 AA; 38722 MW; 635563 CN;

  Query Match           61.5%;  Score 48;  DB 1;  Length 362;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     98 YLAVTNPLRY 107
          ||: :::|:|
Qy      3 YLSTSSSLHY 12
```

```
RESULT    9
ID   US-08-087-772A-2     STANDARD;     PRT;   388 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08087772A
XX
CC    Sequence 2, Application US/08087772A
CC    Patent No. 5691155
CC     GENERAL INFORMATION:
CC       APPLICANT:  Nahmias, Clara
CC       APPLICANT:  Emorine, Jean L.
CC       APPLICANT:  Strosberg, Donny A.
CC       TITLE OF INVENTION:  Nucleotide Sequences Encoding the Murine
CC       TITLE OF INVENTION:  Beta3-Adrenergic Receptor and Their Applications
CC       NUMBER OF SEQUENCES:  17
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Bell, Seltzer, Park & Gibson
CC         STREET:  Post Office Drawer 34009
CC         CITY:  Charlotte
CC         STATE:  No. 5691155th Carolina
CC         COUNTRY:  USA
CC         ZIP:  28234
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/087,772A
CC         FILING DATE:
CC         CLASSIFICATION:  800
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Linker, Raymond O.
CC         REGISTRATION NUMBER:  26,419
CC         REFERENCE/DOCKET NUMBER:  3339-195
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  919-881-3140
CC         TELEFAX:  919-881-3175
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  388 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   388 AA; 41667 MW; 771830 CN;

  Query Match           61.5%;  Score 48;  DB 1;  Length 388;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    133 YLAVTNPLRY 142
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT   10
ID   US-07-916-901-6     STANDARD;     PRT;   400 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/07916901
XX
CC    Sequence 6, Application US/07916901
CC    Patent No. 5364772
CC     GENERAL INFORMATION:
CC       APPLICANT:  Granneman, James G.
CC       APPLICANT:  Lahners, Kristine N.
CC       APPLICANT:  Rao, Donald D.
CC       TITLE OF INVENTION:  @  @3-ADRENERGIC RECEPTOR PROTEIN AND DNA
CC       TITLE OF INVENTION:  ENCODING SAME
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  REISING, ETHINGTON, BARNARD, PERRY &
CC         ADDRESSEE:  MILTON
CC         STREET:  201 W. Big Beaver - Ste. 400; P.O. Box 4390
CC         CITY:  Troy
CC         STATE:  Michigan
CC         COUNTRY:  USA
CC         ZIP:  48099
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/916,901
CC         FILING DATE:  19920720
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Kohn, Kenneth I.
CC         REGISTRATION NUMBER:  30,955
CC         REFERENCE/DOCKET NUMBER:  P-324(WSU)
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (313) 689-3554
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  400 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   400 AA; 43146 MW; 840185 CN;

  Query Match           61.5%;  Score 48;  DB 1;  Length 400;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    133 YLAVTNPLRY 142
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT   11
ID   US-08-351-473B-4     STANDARD;     PRT;   400 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08351473B
XX
CC    Sequence 4, Application US/08351473B
CC    Patent No. 5656440
CC     GENERAL INFORMATION:
CC       APPLICANT:  LENZEN, GERLINDA
CC       APPLICANT:  KAPOOR, ARCHANA
CC       TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES CODING FOR THE
CC       TITLE OF INVENTION:  BOVINE BETA3-ADRENERGIC RECEPTOR AND THEIR APPLICATIONS
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
CC         STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC         CITY:  ARLINGTON
CC         STATE:  VIRGINIA
CC         COUNTRY:  USA
CC         ZIP:  22202
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
```

```
CC         APPLICATION NUMBER:  US/08/351,473B
CC         FILING DATE:  21-FEB-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  93 04670
CC         FILING DATE:  21-APR-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/FR94/00447
CC         FILING DATE:  21-APR-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  OBLON, NORMAN F.
CC         REGISTRATION NUMBER:  24,618
CC         REFERENCE/DOCKET NUMBER:  6639-001-0X PCT
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703) 413-3000
CC         TELEFAX:  (703) 413-2220
CC         TELEX:  248855 OPAT UR
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  400 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   400 AA; 43146 MW; 839778 CN;

  Query Match           61.5%;  Score 48;  DB 1;  Length 400;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches     4;  Conservative     5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    133 YLAVTNPLRY 142
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT   12
ID   US-08-351-473B-5     STANDARD;     PRT;   400 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/08351473B
XX
CC    Sequence 5, Application US/08351473B
CC    Patent No. 5656440
CC     GENERAL INFORMATION:
CC       APPLICANT:  LENZEN, GERLINDA
CC       APPLICANT:  KAPOOR, ARCHANA
CC       TITLE OF INVENTION:  NUCLEOTIDE SEQUENCES CODING FOR THE
CC       TITLE OF INVENTION:  BOVINE BETA3-ADRENERGIC RECEPTOR AND THEIR APPLICATIONS
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
CC         STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC         CITY:  ARLINGTON
CC         STATE:  VIRGINIA
CC         COUNTRY:  USA
CC         ZIP:  22202
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/351,473B
CC         FILING DATE:  21-FEB-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  93 04670
CC         FILING DATE:  21-APR-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/FR94/00447
CC         FILING DATE:  21-APR-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  OBLON, NORMAN F.
CC         REGISTRATION NUMBER:  24,618
CC         REFERENCE/DOCKET NUMBER:  6639-001-0X PCT
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703) 413-3000
CC         TELEFAX:  (703) 413-2220
CC         TELEX:  248855 OPAT UR
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  400 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   400 AA; 43005 MW; 820814 CN;

  Query Match           61.5%;  Score 48;  DB 1;  Length 400;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches     4;  Conservative     5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    133 YLAVTNPLRY 142
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT   13
ID   US-07-783-602C-1     STANDARD;     PRT;   400 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/07783602C
XX
CC    Sequence 1, Application US/07783602C
CC    Patent No. 5418160
CC     GENERAL INFORMATION:
CC       APPLICANT:  J. Craig Venter et al
CC       TITLE OF INVENTION:  A FAT CELL SPECIFIC a-ADRENERGIC
CC       TITLE OF INVENTION:  RECEPTOR
CC       NUMBER OF SEQUENCES:  1
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Lowe, Price, LeBlanc & Becker
CC         STREET:  Suite 300, 99 Canal Center Plaza
CC         CITY:  Alexandria
CC         STATE:  Virginia
CC         COUNTRY:  USA
CC         ZIP:  22314
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  DOS Text File
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/783,602C
CC         FILING DATE:  19911101
CC         CLASSIFICATION:  436
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  J.G. Mullins
CC         REGISTRATION NUMBER:  33073
CC         REFERENCE/DOCKET NUMBER:  717-098
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  703 684 1111
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  400
CC         TYPE:  AMINO ACID
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  Polypeptide
SQ   SEQUENCE   400 AA; 43166 MW; 843002 CN;
```

```
  Query Match           61.5%;  Score 48;  DB 1;  Length 400;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    133 YLAVTNPLRY 142
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT   14
ID   US-08-444-734A-6     STANDARD;     PRT;   402 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08444734A
XX
     Sequence 6, Application US/08444734A
     Patent No. 5610282
     GENERAL INFORMATION:
CC     APPLICANT:  Sibley, David R.
CC     APPLICANT:  Monsma, Frederick J.
CC     APPLICANT:  Mahan, Lawrence C.
CC     APPLICANT:  McVittie, Loris D.
CC     TITLE OF INVENTION:  cDNA encoding the rat D1 dopamine
CC     TITLE OF INVENTION:  receptor linked to adnylyl cyclase activation and
CC     TITLE OF INVENTION:  expression of the receptor protein in plasmid-transfected
CC     TITLE OF INVENTION:  cell lines
CC     NUMBER OF SEQUENCES:  13
CC     CORRESPONDENCE ADDRESS:
CC       ADDRESSEE:  Knobbe, Martens, Olson and Bear
CC       STREET:  620 Newport Center Drive, Sixteenth Floor
CC       CITY:  Newport Beach
CC       STATE:  CA
CC       COUNTRY:  USA
CC       ZIP:  92660
CC     COMPUTER READABLE FORM:
CC       MEDIUM TYPE:  Floppy disk
CC       COMPUTER:  IBM PC compatible
CC       OPERATING SYSTEM:  PC-DOS/MS-DOS
CC       SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC     CURRENT APPLICATION DATA:
CC       APPLICATION NUMBER:  US/08/444,734A
CC       FILING DATE:
CC       CLASSIFICATION:  530
CC     PRIOR APPLICATION DATA:
CC       APPLICATION NUMBER:  US 08/029,917
CC       FILING DATE:  03-MAR-1993
CC     PRIOR APPLICATION DATA:
CC       APPLICATION NUMBER:  US 07/548,714
CC       FILING DATE:  06-JUL-1990
CC     ATTORNEY/AGENT INFORMATION:
CC       NAME:  Altman, Daniel E.
CC       REGISTRATION NUMBER:  34,115
CC       REFERENCE/DOCKET NUMBER:  NIH065.001FW1
CC     TELECOMMUNICATION INFORMATION:
CC       TELEPHONE:  (714) 760-0404
CC       TELEFAX:  (714) 760-9502
CC   INFORMATION FOR SEQ ID NO:  6:
CC     SEQUENCE CHARACTERISTICS:
CC       LENGTH:  402 amino acids
CC       TYPE:  amino acid
CC       STRANDEDNESS:  single
CC       TOPOLOGY:  linear
CC     MOLECULE TYPE:  protein
CC     HYPOTHETICAL:  NO
CC     ANTI-SENSE:  NO
CC     FRAGMENT TYPE:  internal
SQ   SEQUENCE   402 AA; 42931 MW; 794255 CN;

  Query Match           61.5%;  Score 48;  DB 1;  Length 402;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    136 YLAVTNPLRY 145
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT   15
ID   US-08-087-772A-15     STANDARD;     PRT;   402 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 15, Application US/08087772A
XX
CC    Sequence 15, Application US/08087772A
CC    Patent No. 5691155
CC     GENERAL INFORMATION:
CC       APPLICANT:  Nahmias, Clara
CC       APPLICANT:  Emorine, Jean L.
CC       APPLICANT:  Strosberg, Donny A.
CC       TITLE OF INVENTION:  Nucleotide Sequences Encoding the Murine
CC       TITLE OF INVENTION:  Beta3-Adrenergic Receptor and Their Applications
CC       NUMBER OF SEQUENCES:  17
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Bell, Seltzer, Park & Gibson
CC         STREET:  Post Office Drawer 34009
CC         CITY:  Charlotte
CC         STATE:  No. 5691155th Carolina
CC         COUNTRY:  USA
CC         ZIP:  28234
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/087,772A
CC         FILING DATE:
CC         CLASSIFICATION:  800
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Linker, Raymond O.
CC         REGISTRATION NUMBER:  26,419
CC         REFERENCE/DOCKET NUMBER:  3339-195
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  919-881-3140
CC         TELEFAX:  919-881-3175
CC     INFORMATION FOR SEQ ID NO:  15:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  402 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   402 AA; 42931 MW; 794255 CN;

  Query Match           61.5%;  Score 48;  DB 1;  Length 402;
  Best Local Similarity  40.0%;  Pred. No. 1.23e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    136 YLAVTNPLRY 145
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


Search completed: Thu Sep  2 12:44:28 1999
Job time : 7 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

on: Thu Sep 2 12:43:10 1999; MasPar time 2.22 Seconds
153.061 Million cell updates/sec
ular output not generated.

Title: >US-08-599-226-32
Description: (1-12) from US08599226.pep
Perfect Score: 78
Sequence: 1 ASYLSTSSSLHY 12

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 24.890; Variance 28.198; scale 0.883

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	52	66.7	594	1	COX1_PHYPO	CYTOCHROME C OXIDASE P	2.45e+00
2	52	66.7	828	1	YKR6_YEAST	HYPOTHETICAL 93.4 KD P	2.45e+00
3	51	65.4	303	1	YNA6_YEAST	HYPOTHETICAL 34.0 KD T	4.01e+00
4	51	65.4	428	1	B4AR_MELGA	BETA-4C ADRENERGIC REC	4.01e+00
5	51	65.4	543	1	YJF8_YEAST	HYPOTHETICAL 60.8 KD P	4.01e+00
6	50	64.1	328	1	YGJR_ECOLI	HYPOTHETICAL 36.2 KD P	6.49e+00
7	50	64.1	445	1	HPPD_ARATH	4-HYDROXYPHENYLPYRUVAT	6.49e+00
8	50	64.1	1010	1	CLPP_CHLEU	PROBABLE CLPP-LIKE PRO	6.49e+00
9	49	62.8	493	1	PLTP_HUMAN	PHOSPHOLIPID TRANSFER	1.04e+01
10	49	62.8	2410	1	POL1_BAYMJ	GENOME POLYPROTEIN 1 [	1.04e+01
11	49	62.8	2412	1	POL1_BAYMG	GENOME POLYPROTEIN 1 [	1.04e+01
12	48	61.5	158	1	DHSD_BOVIN	SUCCINATE DEHYDROGENAS	1.67e+01
13	48	61.5	159	1	DHSD_HUMAN	SUCCINATE DEHYDROGENAS	1.67e+01
14	48	61.5	351	1	B3AR_CAVPO	BETA-3 ADRENERGIC RECE	1.67e+01
15	48	61.5	400	1	B3AR_MOUSE	BETA-3 ADRENERGIC RECE	1.67e+01
16	48	61.5	400	1	B3AR_RAT	BETA-3 ADRENERGIC RECE	1.67e+01
17	48	61.5	405	1	B3AR_BOVIN	BETA-3 ADRENERGIC RECE	1.67e+01
18	48	61.5	405	1	B3AR_CANFA	BETA-3 ADRENERGIC RECE	1.67e+01
19	48	61.5	408	1	B3AR_HUMAN	BETA-3 ADRENERGIC RECE	1.67e+01
20	48	61.5	418	1	B3AR_MACMU	BETA-3 ADRENERGIC RECE	1.67e+01
21	48	61.5	441	1	SECY_MYCTU	PREPROTEIN TRANSLOCASE	1.67e+01
22	48	61.5	590	1	PYRG_MYCLE	CTP SYNTHASE (EC 6.3.4	1.67e+01
23	48	61.5	963	1	IRE2_RAT	IRON-RESPONSIVE ELEMEN	1.67e+01
24	48	61.5	963	1	IRE2_HUMAN	IRON-RESPONSIVE ELEMEN	1.67e+01
25	48	61.5	1038	1	SOG_DROME	DORSAL-VENTRAL PATTERN	1.67e+01
26	48	61.5	1103	1	CYGD_HUMAN	RETINAL GUANYLYL CYCLA	1.67e+01
27	48	61.5	1255	1	ERB2_HUMAN	ERBB-2 RECEPTOR PROTEI	1.67e+01
28	48	61.5	1469	1	DP27_CAEEL	CHROMOSOME CONDENSATIO	1.67e+01
29	47	60.3	112	1	Y143_MYCGE	HYPOTHETICAL PROTEIN M	2.64e+01
30	47	60.3	291	1	HMP1_BOVIN	PITUITARY-SPECIFIC POS	2.64e+01
31	47	60.3	291	1	HMP1_MOUSE	PITUITARY-SPECIFIC POS	2.64e+01
32	47	60.3	291	1	HMP1_HUMAN	PITUITARY-SPECIFIC POS	2.64e+01
33	47	60.3	291	1	HMP1_PIG	PITUITARY-SPECIFIC POS	2.64e+01
34	47	60.3	291	1	HMP1_SHEEP	PITUITARY-SPECIFIC POS	2.64e+01
35	47	60.3	394	1	BENE_ACICA	BENZOATE MEMBRANE TRAN	2.64e+01
36	47	60.3	434	1	KES1_YEAST	KES1 PROTEIN.	2.64e+01
37	47	60.3	533	1	YADC_SCHPO	HYPOTHETICAL 62.2 KD P	2.64e+01
38	47	60.3	646	1	COAT_ASFB7	MAJOR CAPSID PROTEIN (	2.64e+01
39	47	60.3	684	1	YV18_MYCTU	HYPOTHETICAL 57.3 KD P	2.64e+01
40	47	60.3	1941	1	YRM8_CAEEL	HYPOTHETICAL 216.3 KD	2.64e+01
41	46	59.0	132	1	YW07_MYCTU	HYPOTHETICAL 14.7 KD P	4.14e+01
42	46	59.0	184	1	YBET_ECOLI	HYPOTHETICAL 20.9 KD P	4.14e+01
43	46	59.0	435	1	YT13_CAEEL	HYPOTHETICAL 48.6 KD P	4.14e+01
44	46	59.0	530	1	YA9A_SCHPO	HYPOTHETICAL 54.2 KD S	4.14e+01
45	46	59.0	1844	1	POLR_TYMVC	RNA REPLICASE POLYPROT	4.14e+01

ALIGNMENTS

```
RESULT    1
ID   COX1_PHYPO     STANDARD;      PRT;   594 AA.
AC   Q07434;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1.
OS   PHYSARUM POLYCEPHALUM (SLIME MOLD).
OG   MITOCHONDRION.
OC   EUKARYOTA; MYXOMYCETES; PHYSARIDA; PHYSARUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M3C;
RX   MEDLINE; 94064614.
RA   GOTT J.M., VISOMIRSKI L.M., HUNTER J.L.;
RT   "Substitutional and insertional RNA editing of the cytochrome c
RT   oxidase subunit 1 mRNA of Physarum polycephalum.";
RL   J. BIOL. CHEM. 268:25483-25486(1993).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L14769; E101261; -.
DR   PROSITE; PS00077; COX1; 1.
DR   PFAM; PF00115; COX1; 2.
DR   HSSP; P00396; 1OCC.
KW   OXIDOREDUCTASE; HEME; COPPER; MITOCHONDRION; TRANSMEMBRANE;
KW   RESPIRATORY CHAIN; INNER MEMBRANE.
```

```
FT   METAL        87     87       IRON (HEME A) (PROBABLE).
FT   METAL       266    266       COPPER B (PROBABLE).
FT   METAL       270    270       COPPER B (PROBABLE).
FT   METAL       315    315       COPPER B (PROBABLE).
FT   METAL       316    316       COPPER B (PROBABLE).
FT   METAL       401    401       IRON (HEME A3) (PROBABLE).
FT   METAL       403    403       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   594 AA;  66630 MW;  E9280BA7 CRC32;

  Query Match           66.7%;  Score 52;  DB 1;  Length 594;
  Best Local Similarity  54.5%;  Pred. No. 2.45e+00;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    457 SYLETAGILHF 467
          ||| |:: ||:
Qy      2 SYLSTSSSLHY 12


RESULT    2
     YKR6_YEAST     STANDARD;      PRT;   828 AA.
     P34239;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 93.4 KD PROTEIN IN STE3-GIN10 INTERGENIC REGION.
GN   YKL176C OR YKL642.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94205264.
RA   WIEMANN S., VOSS H., SCHWAGER C., RUPP T., STEGEMANN J.,
RA   ZIMMERMANN J., GROTHUES D., SENSEN C., ERFLE H., HEWITT N.,
RA   BANREVI A., ANSORGE W.;
RT   "Sequencing and analysis of 51.6 kilobases on the left arm of
RT   chromosome XI from Saccharomyces cerevisiae reveals 23 open reading
RT   frames including the FAS1 gene.";
RL   YEAST 9:1343-1348(1993).
RN   [2]
RP   SEQUENCE OF 1-306 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 94378719.
RA   VANDENBOL M., BOLLE P.-A., DION C., PORTETELLE D., HILGER F.;
RT   "Sequencing and analysis of a 20.5 kb DNA segment located on the left
RT   arm of yeast chromosome XI.";
RL   YEAST 10:S25-S33(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74151; G395250; -.
DR   EMBL; Z26878; G407504; -.
DR   EMBL; Z28176; G486310; -.
DR   PIR; S34695; S34695.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   828 AA;  93358 MW;  564E4E57 CRC32;

  Query Match           66.7%;  Score 52;  DB 1;  Length 828;
  Best Local Similarity  66.7%;  Pred. No. 2.45e+00;
  Matches     8;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    483 SSFKSGSSSLHY 494
          :|: | ||||||
Qy      1 ASYLSTSSSLHY 12


RESULT    3
ID   YNA6_YEAST     STANDARD;      PRT;   303 AA.
AC   P41318;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 34.0 KD TRP-ASP REPEATS CONTAINING PROTEIN IN SIS1-MRPL2
DE   INTERGENIC REGION.
GN   YNL006W OR N2005.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE; 95076713.
RA   VERHASSELT P., AERT R., VOET M., VOLCKAERT G.;
RT   "Nucleotide sequence analysis of an 8887 bp region of the left arm of
RT   yeast chromosome XIV, encompassing the centromere sequence.";
RL   YEAST 10:945-951(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   DOIGNON F., CROUZET M.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X77114; G496716; -.
DR   EMBL; Z71282; E239859; -.
DR   PIR; S45123; S45123.
DR   PIR; S45461; S45461.
DR   PROSITE; PS00678; WD_REPEATS; 4.
DR   PFAM; PF00400; G-beta; 5.
KW   HYPOTHETICAL PROTEIN; REPEAT; WD REPEAT.
FT   REPEAT       73    103       WD1.
FT   REPEAT      114    144       WD2.
FT   REPEAT      157    187       WD3.
FT   REPEAT      205    235       WD4.
FT   REPEAT      248    278       WD5.
SQ   SEQUENCE   303 AA;  34034 MW;  ADBFDB90 CRC32;

  Query Match           65.4%;  Score 51;  DB 1;  Length 303;
  Best Local Similarity  58.3%;  Pred. No. 4.01e+00;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    262 SAYLVTASSDHY 273
          ::|| |:|| ||
Qy      1 ASYLSTSSSLHY 12


RESULT    4
ID   B4AR_MELGA     STANDARD;      PRT;   428 AA.
AC   P43141;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   BETA-4C ADRENERGIC RECEPTOR.
GN   ADRB4C.
OS   MELEAGRIS GALLOPAVO (COMMON TURKEY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; MELEAGRIDIDAE; MELEAGRIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95014249.
RA   CHEN X.-H., HARDEN T.K., NICHOLAS R.A.;
RT   "Molecular cloning and characterization of a novel beta-adrenergic
RT   receptor.";
```

```
RL   J. BIOL. CHEM. 269:24810-24819(1994).
CC   -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC       INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC       PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: BROAD TISSUE DISTRIBUTION.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U13977; G556604; -.
DR   EMBL; U13978; G555882; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   HSSP; P07700; 1DEP.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT   DOMAIN        1     25       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     26     49       1 (POTENTIAL).
FT   DOMAIN       50     58       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     59     77       2 (POTENTIAL).
FT   DOMAIN       78     97       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     98    119       3 (POTENTIAL).
FT   DOMAIN      120    141       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    142    164       4 (POTENTIAL).
FT   DOMAIN      165    189       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    190    211       5 (POTENTIAL).
FT   DOMAIN      212    261       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    262    283       6 (POTENTIAL).
FT   DOMAIN      284    294       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    295    315       7 (POTENTIAL).
FT   DOMAIN      316    428       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      8      8       POTENTIAL.
FT   CARBOHYD     13     13       POTENTIAL.
FT   DISULFID     96    175       BY SIMILARITY.
FT   LIPID       329    329       PALMITATE (BY SIMILARITY).
SQ   SEQUENCE   428 AA;  47398 MW;  8B794F0C CRC32;

  Query Match           65.4%;  Score 51;  DB 1;  Length 428;
  Best Local Similarity  40.0%;  Pred. No. 4.01e+00;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    122 YLAITAPLQY 131
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT    5
ID   YJF8_YEAST     STANDARD;      PRT;   543 AA.
AC   P47041;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 60.8 KD PROTEIN IN BTN1-PEP8 INTERGENIC REGION.
GN   YJL058C OR J1141.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   POHL T.M., ALJINOVIC G.;
RL   SUBMITTED (SEP-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: STRONG, TO YEAST YBR270C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49333; G1008201; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   543 AA;  60840 MW;  8233FB93 CRC32;

  Query Match           65.4%;  Score 51;  DB 1;  Length 543;
  Best Local Similarity  63.6%;  Pred. No. 4.01e+00;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    182 ASFLTDSSTLH 192
          ||:|: ||:||
Qy      1 ASYLSTSSSLH 11


RESULT    6
ID   YGJR_ECOLI     STANDARD;      PRT;   328 AA.
AC   P42599; P42600; P76661;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 36.2 KD PROTEIN IN EBGC-UXAA INTERGENIC REGION.
GN   YGJR.
OS   ESCHERICHIA COLI.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ESCHERICHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   SCIENCE 277:1453-1474(1997).
CC   -!- SIMILARITY: BELONGS TO THE GFO/IDH/MOCA FAMILY. STRONG, TO
CC       B.SUBTILIS YULF.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18997; G606025; ALT_FRAME.
DR   EMBL; U18997; G606026; ALT_FRAME.
DR   EMBL; AE000390; G2367190; ALT_INIT.
DR   ECOGENE; EG12729; YGJR.
KW   HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT   TRANSMEM    177    193       POTENTIAL.
SQ   SEQUENCE   328 AA;  36214 MW;  236E023A CRC32;

  Query Match           64.1%;  Score 50;  DB 1;  Length 328;
  Best Local Similarity  41.7%;  Pred. No. 6.49e+00;
  Matches     5;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db     66 AVYIASPNSLHF 77
          | |:::::|||:
Qy      1 ASYLSTSSSLHY 12


RESULT    7
ID   HPPD_ARATH     STANDARD;      PRT;   445 AA.
AC   P93836; O04330;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
```

```
DE   4-HYDROXYPHENYLPYRUVATE DIOXYGENASE (EC 1.13.11.27) (4HPPD) (HPD).
GN   HPD.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. WASSILEWSKIJA;
RA   BARTLEY G.E., MAXWELL C.A., WITTENBACH V.A., SCOLNIK P.A.;
RT   "Cloning of an Arabidopsis thaliana cDNA for p-hydroxyphenylpyruvate
RT   dioxygenase.";
RL   (IN) PLANT GENE REGISTER PGR97-065.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   NORRIS S.R., DELLAPENNA D.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   HSIEH T.-F., RODGERS M., MATRINGE M.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: 4-HYDROXYPHENYLPYRUVATE + O(2) =
CC       HOMOGENTISATE + CO(2).
CC   -!- COFACTOR: IRON (BY SIMILARITY).
CC   -!- PATHWAY: PARTICIPATES IN THE CATABOLISM OF TYROSINE. IN THE
CC       PATHWAY FOR BIOSYNTHESIS OF PRENYLQUINONES.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE 4HPPD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U89267; G2392518; -.
DR   EMBL; AF000228; G2145039; -.
DR   EMBL; AF047834; G3098559; -.
KW   OXIDOREDUCTASE; DIOXYGENASE; IRON.
SQ   SEQUENCE   445 AA;  48816 MW;  B88784CA CRC32;

  Query Match           64.1%;  Score 50;  DB 1;  Length 445;
  Best Local Similarity  58.3%;  Pred. No. 6.49e+00;
  Matches    7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

      86 ASYLLTSGDLRF 97
         |||| ||: |::
Qy     1 ASYLSTSSSLHY 12


RESULT    8
ID   CLPP_CHLEU     STANDARD;      PRT;  1010 AA.
AC   P42379;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PROBABLE CLPP-LIKE PROTEASE (EC 3.4.21.92) (ENDOPEPTIDASE CLP).
GN   CLPP.
OS   CHLAMYDOMONAS EUGAMETOS.
OG   CHLOROPLAST.
OC   EUKARYOTA; VIRIDIPLANTAE; CHLOROPHYTA; CHLOROPHYCEAE; VOLVOCALES;
OC   CHLAMYDOMONADACEAE; CHLAMYDOMONAS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94329067.
RA   HUANG C., WANG S., CHEN L., LEMIEUX C., OTIS C., TURMEL M.,
RA   LIU X.-Q.;
RT   "The Chlamydomonas chloroplast clpP gene contains translated large
RT   insertion sequences and is essential for cell growth.";
RL   MOL. GEN. GENET. 244:151-159(1994).
RN   [2]
RP   PROTEIN SPLICING.
RX   MEDLINE; 97277324.
RA   WANG S., LIU X.-Q.;
RT   "Identification of an unusual intein in chloroplast ClpP protease of
RT   Chlamydomonas eugametos.";
RL   J. BIOL. CHEM. 272:11869-11873(1997).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS TO SMALL PEPTIDES IN
CC       THE PRESENCE OF ATP AND MAGNESIUM. ALPHA-CASEIN IS THE USUAL TEST
CC       SUBSTRATE. IN THE ABSENCE OF ATP, ONLY OLIGOPEPTIDES SHORTER THAN
CC       FIVE RESIDUES ARE CLEAVED (SUCH AS SUCCINYL-LEU-TYR-|-NHMEC; AND
CC       LEU-TYR-LEU-|-TYR-TRP, IN WHICH THE CLEAVAGE OF THE -TYR-|-LEU-
CC       AND -TYR-|-TRP- BOND ALSO OCCURS).
CC   -!- DOMAIN: THIS GENE CONTAINS TWO LARGE INSERTION SEQUENCES (IS1 AND
CC       CEU CLPP INTEIN) THAT DIVIDE THE CLPP GENE INTO THREE SEQUENCE
CC       DOMAINS. EACH INSERTION SEQUENCE FORMS A CONTINUOUS OPEN READING
CC       FRAME WITH ITS UPSTREAM AND DOWNSTREAM SEQUENCE DOMAINS.
CC   -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC       A POST-TRANSLATIONAL EXCISION OF THE THE INTERVENING REGION
CC       (INTEIN) FOLLOWED BY PEPTIDE LIGATION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S14; ALSO KNOWN AS CLPP
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29402; G575471; -.
DR   PROSITE; PS00381; CLP_PROTEASE_SER; 1.
DR   PROSITE; PS00382; CLP_PROTEASE_HIS; 1.
DR   PROSITE; PS00881; PROTEIN_SPLICING; FALSE_NEG.
DR   PFAM; PF00574; CLP_protease; 3.
DR   MENDEL; 2294; CHLeu;clpP;1.
KW   HYDROLASE; SERINE PROTEASE; CHLOROPLAST; PROTEIN SPLICING.
FT   CHAIN         1    447       CLPP, 1ST PART.
FT   CHAIN       448    903       CEU CLPP INTEIN (INSERTION IS2).
FT   CHAIN       904   1010       CLPP, 2ND PART.
FT   DOMAIN       60    377       INSERTION IS1.
FT   ACT_SITE    419    419       BY SIMILARITY.
FT   ACT_SITE    444    444       BY SIMILARITY.
SQ   SEQUENCE   1010 AA;  114551 MW;  D08E3DA6 CRC32;

  Query Match           64.1%;  Score 50;  DB 1;  Length 1010;
  Best Local Similarity  40.0%;  Pred. No. 6.49e+00;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    951 YLTATETIHY 960
          ||::: ::||
Qy      3 YLSTSSSLHY 12


RESULT    9
ID   PLTP_HUMAN     STANDARD;      PRT;   493 AA.
AC   P55058;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PHOSPHOLIPID TRANSFER PROTEIN PRECURSOR (LIPID TRANSFER PROTEIN II).
GN   PLTP.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 18-27 AND 163-184.
RC   TISSUE=UMBILICAL VEIN ENDOTHELIAL CELLS;
RX   MEDLINE; 94179366.
RA   DAY J.R., ALBERS J.J., LOFTON-DAY C.E., GILBERT T.L., CHING A.F.T.,
RA   GRANT F.J., O'HARA P.J., MARCOVINA S.M., ADOLPHSON J.L.;
```

```
RT   "Complete cDNA encoding human phospholipid transfer protein from
RT   human endothelial cells.";
RL   J. BIOL. CHEM. 269:9388-9391(1994).
CC   -!- FUNCTION: CONVERTS HDL INTO LARGER AND SMALLER PARTICLES. MAY PLAY
CC       A KEY ROLE IN EXTRACELLULAR PHOSPHOLIPID TRANSPORT AND MODULATION
CC       OF HDL PARTICLES.
CC   -!- TISSUE SPECIFICITY: WIDE TISSUE DISTRIBUTION. PLACENTA > PANCREAS
CC       > LUNG > KIDNEY > HEART > LIVER > SKELETAL MUSCLE > BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE BPI/CETP/LBP/PLTP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L26232; G468326; -.
DR   MIM; 172425; -.
DR   PROSITE; PS00400; LBP_BPI_CETP; 1.
DR   PFAM; PF01273; LBP_BPI_CETP; 1.
KW   LIPID TRANSPORT; GLYCOPROTEIN; SIGNAL.
FT   SIGNAL        1     17
FT   CHAIN        18    493       PHOSPHOLIPID TRANSFER PROTEIN.
FT   CARBOHYD     64     64       POTENTIAL.
FT   CARBOHYD     94     94       POTENTIAL.
FT   CARBOHYD    117    117       POTENTIAL.
FT   CARBOHYD    143    143       POTENTIAL.
FT   CARBOHYD    245    245       POTENTIAL.
FT   CARBOHYD    398    398       POTENTIAL.
SQ   SEQUENCE   493 AA;  54739 MW;  C178FB86 CRC32;

  Query Match           62.8%;  Score 49;  DB 1;  Length 493;
  Best Local Similarity  33.3%;  Pred. No. 1.04e+01;
  Matches     4;  Conservative    6;  Mismatches   2;  Indels   0;  Gaps   0;

Db    453 AGFLTIGADLHF 464
          |::|: :: ||:
Qy      1 ASYLSTSSSLHY 12


RESULT   10
ID   POL1_BAYMJ     STANDARD;      PRT;  2410 AA.
AC   Q01206;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   GENOME POLYPROTEIN 1 [CONTAINS: CYTOPLASMIC INCLUSION PROTEIN (CI);
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48); COAT PROTEIN (CP)].
OS   BARLEY YELLOW MOSAIC VIRUS (JAPANESE STRAIN II-1) (BAYMV).
OC   VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC   BYMOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91108371.
RA   KASHIWAZAKI S., MINOBE Y., OMURA T., HIBINO H.;
RT   "Nucleotide sequence of barley yellow mosaic virus RNA 1: a close
RT   evolutionary relationship with potyviruses.";
RL   J. GEN. VIROL. 71:2781-2790(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90063549.
RA   KASHIWAZAKI S., HAYANO Y., MINOBE Y., OMURA T., HIBINO H.,
RA   TSUCHIZAKI T.;
RT   "Nucleotide sequence of the capsid protein gene of barley yellow
RT   mosaic virus.";
RL   J. GEN. VIROL. 70:3015-3023(1989).
CC   -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC       MAY BE INVOLVED IN REPLICATION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D01091; D1001339; -.
DR   PIR; JQ1948; JQ1948.
DR   PFAM; PF00271; helicase_C; 1.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00767; Poty_coat; 1.
DR   PFAM; PF00863; Peptidase_C4; 1.
KW   TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; COAT PROTEIN; POLYPROTEIN;
KW   HELICASE; ATP-BINDING.
FT   CHAIN         ?      ?       CYTOPLASMIC INCLUSION PROTEIN.
FT   CHAIN         ?   2113       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN      2114   2410       COAT PROTEIN.
FT   NP_BIND     487    494       ATP (POTENTIAL).
SQ   SEQUENCE   2410 AA;  270767 MW;  FFCF43DA CRC32;

  Query Match           62.8%;  Score 49;  DB 1;  Length 2410;
  Best Local Similarity  58.3%;  Pred. No. 1.04e+01;
  Matches     7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db    883 ASYFSLHQALHY 894
          |||:|   :|||
Qy      1 ASYLSTSSSLHY 12


RESULT   11
ID   POL1_BAYMG     STANDARD;      PRT;  2412 AA.
AC   Q04574;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   GENOME POLYPROTEIN 1 [CONTAINS: CYTOPLASMIC INCLUSION PROTEIN (CI);
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48); COAT PROTEIN (CP)].
OS   BARLEY YELLOW MOSAIC VIRUS (GERMAN ISOLATE) (BAYMV).
OC   VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC   BYMOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92268893.
RA   PEERENBOOM E., PROELS M., SCHELL J., STEINBISS H.-H., DAVIDSON A.D.;
RT   "The complete nucleotide sequence of RNA 1 of a German isolate of
RT   barley yellow mosaic virus and its comparison with a Japanese
RT   isolate.";
RL   J. GEN. VIROL. 73:1303-1308(1992).
CC   -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC       MAY BE INVOLVED IN REPLICATION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69757; G58679; -.
DR   PIR; JQ1537; JQ1537.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00767; Poty_coat; 1.
DR   PFAM; PF00863; Peptidase_C4; 1.
KW   TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; COAT PROTEIN; POLYPROTEIN;
KW   HELICASE; ATP-BINDING.
FT   CHAIN         ?      ?       CYTOPLASMIC INCLUSION PROTEIN.
FT   CHAIN         ?   2115       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN      2116   2412       COAT PROTEIN.
FT   NP_BIND     487    494       ATP (POTENTIAL).
SQ   SEQUENCE   2412 AA;  270873 MW;  6BC96278 CRC32;

  Query Match           62.8%;  Score 49;  DB 1;  Length 2412;
```

```
  Best Local Similarity  58.3%;  Pred. No. 1.04e+01;
  Matches     7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db    883 ASYFSLHQALHY 894
          |||:|   :|||
Qy      1 ASYLSTSSSLHY 12


RESULT   12
ID   DHSD_BOVIN     STANDARD;      PRT;   158 AA.
AC   Q95123;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT
DE   PRECURSOR (CYBS) (SUCCINATE-UBIQUINONE REDUCTASE MEMBRANE ANCHOR
DE   SUBUNIT) (QPS3).
GN   SDHD OR SDH4.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RA   SHENOY S.K., YU L., YU C.A.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBUNIT: COMPOSED OF A 27 KD IRON PROTEIN (IP), A 70 KD
CC       FLAVOPROTEIN (FP) AND A CYTOCHROME B COMPOSED OF TWO INTEGRAL
CC       MEMBRANE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U50987; G1575011; -.
KW   TRICARBOXYLIC ACID CYCLE; ELECTRON TRANSPORT; HEME; TRANSMEMBRANE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   TRANSIT       1     55       MITOCHONDRION (POTENTIAL).
FT   CHAIN        56    158       SUCCINATE DEHYDROGENASE [UBIQUINONE]
FT                                 CYTOCHROME B SMALL SUBUNIT.
FT   TRANSMEM     70     90       POTENTIAL.
FT   TRANSMEM    125    141       POTENTIAL.
FT   SEQUENCE   158 AA;  17096 MW;  703D5238 CRC32;

  Query Match           61.5%;  Score 48;  DB 1;  Length 158;
  Best Local Similarity  41.7%;  Pred. No. 1.67e+01;
  Matches     5;  Conservative    4;  Mismatches   3;  Indels   0;  Gaps   0;

Db     81 AAYLNPCSAMDY 92
          |:||:  |:: |
Qy      1 ASYLSTSSSLHY 12


RESULT   13
ID   DHSD_HUMAN     STANDARD;      PRT;   159 AA.
AC   O14521;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT
DE   PRECURSOR (CYBS) (SUCCINATE-UBIQUINONE REDUCTASE MEMBRANE ANCHOR
DE   SUBUNIT).
GN   SDHD OR SDH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 98194224.
RA   HIRAWAKE H., TANIWAKI M., KIJIMA S., KITA K.;
RT   "Cytochrome b in human complex II (succinate-ubiquinone
RT   oxidoreductase): cDNA cloning of the components in liver mitochondria
RT   and chromosome assignment of the genes for the large (SDHC) and small
RT   (SDHD) subunits to 1q21 and 11q23.";
RL   CYTOGENET. CELL GENET. 79:132-138(1997).
CC   -!- SUBUNIT: COMPOSED OF A 27 KD IRON PROTEIN (IP), A 70 KD
CC       FLAVOPROTEIN (FP) AND A CYTOCHROME B COMPOSED OF TWO INTEGRAL
CC       MEMBRANE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB006202; D1022913; -.
DR   MIM; 602690; -.
KW   TRICARBOXYLIC ACID CYCLE; ELECTRON TRANSPORT; HEME; TRANSMEMBRANE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   TRANSIT       1     56       MITOCHONDRION (POTENTIAL).
FT   CHAIN        57    159       SUCCINATE DEHYDROGENASE [UBIQUINONE]
FT                                 CYTOCHROME B SMALL SUBUNIT.
FT   TRANSMEM     71     91       POTENTIAL.
FT   TRANSMEM    126    142       POTENTIAL.
SQ   SEQUENCE   159 AA;  17043 MW;  F4221825 CRC32;

  Query Match           61.5%;  Score 48;  DB 1;  Length 159;
  Best Local Similarity  41.7%;  Pred. No. 1.67e+01;
  Matches     5;  Conservative    4;  Mismatches   3;  Indels   0;  Gaps   0;

Db     82 AAYLNPCSAMDY 93
          |:||:  |:: |
Qy      1 ASYLSTSSSLHY 12


RESULT   14
ID   B3AR_CAVPO     STANDARD;      PRT;   351 AA.
AC   Q60483;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   BETA-3 ADRENERGIC RECEPTOR (FRAGMENT).
GN   ADRB3.
OS   CAVIA PORCELLUS (GUINEA PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; HYSTRICOGNATHI; CAVIIDAE; CAVIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 97151378.
RA   ATGIE C., TAVERNIER G., D'ALLAIRE F., BENGTSSON T., MARTI L.,
RA   CARPENE C., LAFONTAN M., BUKOWIECKI L.J., LANGIN D.;
RT   "Beta 3-adrenoceptor in guinea pig brown and white adipocytes: low
RT   expression and lack of function.";
RL   AM. J. PHYSIOL. 271:R1729-R1738(1996).
CC   -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC       INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC       PROTEINS. BETA-3 IS INVOLVED IN THE REGULATION OF LIPOLYSIS AND
CC       THERMOGENESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: WHITE AND BROWN ADIPOSE TISSUES.
CC   -!- THE GUINEA PIG DIFFERS FROM OTHER RODENTS BY AN ABSENCE OF BETA-3
CC       ADRENERGIC EFFECTS AND BY LOW EXPRESSION IN BROWN AND WHITE
CC       ADIPOSE TISSUES. IT IS CLOSER TO HUMAN OR PRIMATE THAN RODENT
CC       BETA-3.
```

```
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U51098; G1256416; -.
DR   GCRDB; GCR_1166; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   MULTIGENE FAMILY.
FT   DOMAIN        1     36       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     37     60       1 (POTENTIAL).
FT   DOMAIN       61     69       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     70     88       2 (POTENTIAL).
FT   DOMAIN       89    108       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    109    130       3 (POTENTIAL).
FT   DOMAIN      131    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    175       4 (POTENTIAL).
FT   DOMAIN      176    200       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    201    222       5 (POTENTIAL).
FT   DOMAIN      223    290       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    291    312       6 (POTENTIAL).
FT   DOMAIN      313    324       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    325    345       7 (POTENTIAL).
FT   DOMAIN      346   >351       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    107    186       BY SIMILARITY.
FT   CARBOHYD      8      8       POTENTIAL.
FT   CARBOHYD     26     26       POTENTIAL.
FT   NON_TER     351    351
SQ   SEQUENCE   351 AA;  37364 MW;  72E17433 CRC32;

  Query Match           61.5%;  Score 48;  DB 1;  Length 351;
  Best Local Similarity  40.0%;  Pred. No. 1.67e+01;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    133 YLAVTNPLRY 142
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT   15
ID   B3AR_MOUSE     STANDARD;      PRT;   400 AA.
AC   P25962;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BETA-3 ADRENERGIC RECEPTOR.
GN   ADRB3 OR ADRB3R.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS;
RX   MEDLINE; 92037534.
RA   NAHMIAS C., BLIN N., ELALOUF J.M., MATTEI M.-G., STROSBERG A.D.,
RA   EMORINE L.J.;
RT   "Molecular characterization of the mouse beta 3-adrenergic receptor:
RT   relationship with the atypical receptor of adipocytes.";
RL   EMBO J. 10:3721-3727(1991).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RX   MEDLINE; 93279311.
RA   VAN SPRONSEN A., NAHMIAS C., KRIEF S., BRIEND-SUTREN M.-M.,
RA   STROSBERG A.D., EMORINE L.J.;
RT   "The promoter and intron/exon structure of the human and mouse beta
RT   3-adrenergic-receptor genes.";
RL   EUR. J. BIOCHEM. 213:1117-1124(1993).
RN   [3]
RP   SEQUENCE OF 378-400 FROM N.A.
RC   TISSUE=ADIPOSE TISSUE;
RX   MEDLINE; 93125503.
RA   GRANNEMAN J.G., LAHNERS K.N., RAO D.D.;
RT   "Rodent and human beta 3-adrenergic receptor genes contain an intron
RT   within the protein-coding block.";
RL   MOL. PHARMACOL. 42:964-970(1992).
CC   -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC       INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC       PROTEINS. BETA-3 IS INVOLVED IN THE REGULATION OF LIPOLYSIS AND
CC       THERMOGENESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: WHITE AND BROWN ADIPOSE TISSUES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72862; G298113; -.
DR   EMBL; X60438; G50110; -.
DR   EMBL; S53290; G263087; -.
DR   PIR; S18548; S18548.
DR   PIR; S32804; S32804.
DR   GCRDB; GCR_0253; -.
DR   GCRDB; GCR_0551; -.
DR   GCRDB; GCR_0708; -.
DR   MGD; MGI:87939; ADRB3.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   HSSP; P07700; 1DEP.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT   DOMAIN        1     36       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     37     60       1 (POTENTIAL).
FT   DOMAIN       61     69       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     70     88       2 (POTENTIAL).
FT   DOMAIN       89    108       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    109    130       3 (POTENTIAL).
FT   DOMAIN      131    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    175       4 (POTENTIAL).
FT   DOMAIN      176    200       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    201    222       5 (POTENTIAL).
FT   DOMAIN      223    289       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    290    311       6 (POTENTIAL).
FT   DOMAIN      312    323       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    324    344       7 (POTENTIAL).
FT   DOMAIN      345    400       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      8      8       POTENTIAL.
FT   CARBOHYD     26     26       POTENTIAL.
FT   DISULFID    107    186       BY SIMILARITY.
FT   LIPID       358    358       PALMITATE (BY SIMILARITY).
SQ   SEQUENCE   400 AA;  43006 MW;  474A96AC CRC32;

  Query Match           61.5%;  Score 48;  DB 1;  Length 400;
  Best Local Similarity  40.0%;  Pred. No. 1.67e+01;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    133 YLAVTNPLRY 142
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


Search completed: Thu Sep  2 12:43:17 1999
Job time : 7 secs.
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:42:33 1999; MasPar time 3.14 Seconds
152.929 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-32
Description: (1-12) from US08599226.pep
Perfect Score: 78
Sequence: 1 ASYLSTSSSLHY 12

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.334; Variance 30.685; scale 0.793

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	59	75.6	989	2	I56333	apolipoprotein B - ra	2.24e-01
2	52	66.7	296	2	S45336	finger protein, Sp1/e	6.21e+00
3	52	66.7	828	2	S34695	hypothetical protein	6.21e+00
4	51	65.4	303	2	S45461	hypothetical protein	9.75e+00
5	51	65.4	428	2	A55044	beta-4C-adrenergic re	9.75e+00
6	51	65.4	543	2	S56830	probable purine nucle	9.75e+00
7	51	65.4	2150	2	S71629	sensory transduction	9.75e+00
8	50	64.1	236	2	S46280	clpP protein - Chlamy	1.52e+01
9	50	64.1	253	2	A71648	pseudouridylate synth	1.52e+01
10	50	64.1	334	2	D65097	ygjR protein - Escher	1.52e+01
11	49	62.8	135	2	S24320	Ig kappa chain precur	2.36e+01
12	49	62.8	493	2	A53533	phospholipid transfer	2.36e+01
13	49	62.8	503	2	F64713	protein-export membra	2.36e+01
14	49	62.8	526	2	D71805	protein-export membra	2.36e+01
15	49	62.8	968	2	T00353	hypothetical protein	2.36e+01
16	49	62.8	1392	2	T01908	hypothetical protein	2.36e+01
17	49	62.8	2410	1	JQ1948	genome polyprotein 1	2.36e+01
18	49	62.8	2412	1	JQ1537	genome polyprotein 1	2.36e+01
19	48	61.5	400	2	S32804	beta-3-adrenergic rec	3.63e+01
20	48	61.5	400	2	A41679	beta-3-adrenergic rec	3.63e+01
21	48	61.5	400	2	A53281	beta 3-adrenergic rec	3.63e+01
22	48	61.5	405	2	S65459	beta-3-adrenergic rec	3.63e+01
23	48	61.5	408	1	QRHUBE	beta-3-adrenergic rec	3.63e+01
24	48	61.5	414	1	QRHUB3	beta-3-adrenergic rec	3.63e+01
25	48	61.5	418	2	G02953	beta-3-adrenergic rec	3.63e+01
26	48	61.5	441	2	G70822	probable secY protein	3.63e+01
27	48	61.5	590	2	S72961	CTP synthase (EC 6.3.	3.63e+01
28	48	61.5	826	2	B36203	iron-responsive eleme	3.63e+01
29	48	61.5	952	2	B57238	iron-responsive eleme	3.63e+01
30	48	61.5	963	2	A57238	iron-responsive eleme	3.63e+01
31	48	61.5	1255	1	A24571	protein-tyrosine kina	3.63e+01
32	48	61.5	1469	2	A55095	chromosome condensati	3.63e+01
33	47	60.3	105	2	S69755	hypothetical protein	5.56e+01
34	47	60.3	112	2	H64215	hypothetical protein	5.56e+01
35	47	60.3	179	2	S65534	light-harvesting chlo	5.56e+01
36	47	60.3	265	2	S38380	Hrox1 protein - Calif	5.56e+01
37	47	60.3	291	1	S18718	transcription factor	5.56e+01
38	47	60.3	291	1	A31305	transcription factor	5.56e+01
39	47	60.3	325	2	S75747	hypothetical protein	5.56e+01
40	47	60.3	394	2	S23481	benE protein - Acinet	5.56e+01
41	47	60.3	434	2	S42676	KES1 protein - yeast	5.56e+01
42	47	60.3	469	2	D70048	ABC transporter (amin	5.56e+01
43	47	60.3	533	2	S62489	hypothetical protein	5.56e+01
44	47	60.3	684	2	G70744	hypothetical protein	5.56e+01
45	47	60.3	1746	2	S19694	tenascin precursor -	5.56e+01

ALIGNMENTS

RESULT 1
ENTRY I56333 #type fragment
TITLE apolipoprotein B - rat (fragment)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 23-Feb-1997
ACCESSIONS I56333
REFERENCE I56333
#authors Reuben, M.A.; Svenson, K.L.; Doolittle, M.H.; Johnson, D.F.; Lusis, A.J.; Elovson, J.
#journal J. Lipid Res. (1988) 29:1337-1347
#title Biosynthetic relationships between three rat apolipoprotein B peptides.
#cross-references MUID:89176719
#accession I56333
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-989 ##label RES
##cross-references GB:M27440; NID:g623548; PID:g623549
GENETICS
#gene apoB
CLASSIFICATION #superfamily apolipoprotein B
SUMMARY #length 989 #checksum 1918

Query Match 75.6%; Score 59; DB 2; Length 989;
Best Local Similarity 70.0%; Pred. No. 2.24e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Db    175 YLQASTSLHY 184
          || :|:|||| 
Qy      3 YLSTSSSLHY 12
```

RESULT 2
ENTRY S45336 #type complete
TITLE finger protein, Sp1/egr-like - fruit fly (Drosophila sp.)
ORGANISM #formal_name Drosophila sp.
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 01-Aug-1997
ACCESSIONS S45336
REFERENCE S45336
#authors Broenner, G.; Chu-LaGraff, Q.; Doe, C.Q.; Cohen, B.; Weigel, D.; Taubert, H.; Jaeckle, H.
#journal Nature (1994) 369:664-668
#title Sp1/egr-like zinc-finger protein required for endoderm specification and germ-layer formation in Drosophila.
#accession S45336

```
     ##status        preliminary
     ##molecule_type DNA
     ##residues      1-296 ##label BRO
SUMMARY          #length 296  #molecular-weight 33598  #checksum 2521

  Query Match           66.7%;  Score 52;  DB 2;  Length 296;
  Best Local Similarity  50.0%;  Pred. No. 6.21e+00;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    127 APFLSAASDLYY 138
          |::||::| | |
Qy      1 ASYLSTSSSLHY 12


RESULT    3
ENTRY            S34695     #type complete
TITLE            hypothetical protein YKL176c - yeast (Saccharomyces
                   cerevisiae)
ALTERNATE_NAMES  hypothetical protein YKL642
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                   21-Nov-1997
ACCESSIONS       S34695; S38008; S38006; S44583; S38403
REFERENCE        S34679
   #authors      Wieman, S.; Voss, H.; Schwagaer, C.; Rupp, T.; Stegemann, J.;
                   Zimmermann, J.; Grothues, D.; Sensen, C.; Erfle, H.;
                   Hewitt, N.; Banrevi, A.; Ansorge, W.
   #submission   submitted to the EMBL Data Library, July 1993
   #description  Sequencing and analysis of 51.5 kilobases on the left arm of
                   chromosome XI from Saccharomyces cerevisiae reveals 23 open
                   reading frames including the FAS1 gene.
   #accession    S34695
      ##molecule_type DNA
      ##residues      1-828 ##label WIE
      ##cross-references EMBL:X74151; NID:g450365; PID:g395250
      ##experimental_source strain S288C
REFERENCE        S37825
   #authors      Wiemann, S.; Voss, H.; Schwager, C.; Rupp, T.; Grothues, D.;
                   Sensen, C.; Stegemann, J.; Zimmermann, J.; Erfle, H.;
                   Hewitt, N.; Ansorge, W.
   #submission   submitted to the Protein Sequence Database, March 1994
   #accession    S38008
      ##molecule_type DNA
      ##residues      1-828 ##label WI2
      ##cross-references EMBL:Z28176; NID:g486309; PID:g486310; MIPS:YKL176c
      ##experimental_source strain S288C
REFERENCE        S37976
   #authors      Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger,
                   F.
   #submission   submitted to the Protein Sequence Database, March 1994
   #accession    S38006
      ##molecule_type DNA
      ##residues      1-306 ##label VAN
      ##cross-references EMBL:Z28176; MIPS:YKL176c
      ##experimental_source strain S288C
REFERENCE        S44583
   #authors      Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger,
                   F.
   #journal      Yeast (1994) 10:25-33
   #title        Sequencing and analysis of a 20.5 kb DNA segment located on
                   the left arm of yeast chromosome XI.
   #accession    S44583
      ##status        translation not shown
      ##molecule_type DNA
      ##residues      1-306 ##label VA2
      ##cross-references EMBL:Z26878; NID:g407503; PID:g407504
      ##experimental_source strain S288C
GENETICS
   #map_position 11L
SUMMARY          #length 828  #molecular-weight 93358  #checksum 6781

  Query Match           66.7%;  Score 52;  DB 2;  Length 828;
  Best Local Similarity  66.7%;  Pred. No. 6.21e+00;
  Matches     8;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    483 SSFKSGSSSLHY 494
          :|: | ||||||
Qy      1 ASYLSTSSSLHY 12


RESULT    4
ENTRY            S45461     #type complete
TITLE            hypothetical protein YNL006w - yeast (Saccharomyces
                   cerevisiae)
ALTERNATE_NAMES  hypothetical protein N2005
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             13-Jan-1995 #sequence_revision 27-Jan-1995 #text_change
                   17-Mar-1999
ACCESSIONS       S45461; S62915; S62917; S45123
REFERENCE        S45456
   #authors      Verhasselt, P.; Aert, R.; Voet, M.; Volckaert, G.
   #journal      Yeast (1994) 10:945-951
   #title        Nucleotide sequence analysis of an 8887 bp region of the left
                   arm of yeast chromosome XIV, encompassing the centromere
                   sequence.
   #cross-references MUID:95076713
   #accession    S45461
      ##status        translation not shown
      ##molecule_type DNA
      ##residues      1-303 ##label VER
      ##cross-references EMBL:X77114; NID:g496710; PID:g496716
REFERENCE        S62910
   #authors      Aert, R.; Verhasselt, P.; Voet, M.; Volckaert, G.
   #submission   submitted to the Protein Sequence Database, April 1996
   #accession    S62915
      ##molecule_type DNA
      ##residues      1-303 ##label AER
      ##cross-references EMBL:Z71282; NID:g1301821; PID:e239859; PID:g1301822;
                        MIPS:YNL006w
      ##experimental_source strain S288C
REFERENCE        S62916
   #authors      Doignon, F.; Crouzet, M.
   #submission   submitted to the Protein Sequence Database, April 1996
   #accession    S62917
      ##molecule_type DNA
      ##residues      1-303 ##label DOI
      ##cross-references EMBL:Z71282; NID:g1301821; PID:e239859; PID:g1301822;
                        MIPS:YNL006w
      ##experimental_source strain S288C
GENETICS
   #map_position 14L
CLASSIFICATION   #superfamily WD repeat homology
FEATURE
   71-104             #domain WD repeat homology #label WD2\
   112-145            #domain WD repeat homology #label WD3\
   203-236            #domain WD repeat homology #label WD5\
   246-279            #domain WD repeat homology #label WD6
SUMMARY          #length 303  #molecular-weight 34034  #checksum 537

  Query Match           65.4%;  Score 51;  DB 2;  Length 303;
  Best Local Similarity  58.3%;  Pred. No. 9.75e+00;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    262 SAYLVTASSDHY 273
          ::|| |:|| ||
Qy      1 ASYLSTSSSLHY 12


RESULT    5
ENTRY            A55044     #type complete
TITLE            beta-4C-adrenergic receptor - turkey
ORGANISM         #formal_name Meleagris gallopavo #common_name common turkey
DATE             18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change
                   08-Sep-1997
ACCESSIONS       A55044
REFERENCE        A55044
```

```
   #authors      Chen, X.; Harden, T.K.; Nicholas, R.A.
   #journal      J. Biol. Chem. (1994) 269:24810-24819
   #title        Molecular cloning and characterization of a novel
                   beta-adrenergic receptor.
   #accession    A55044
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-428 ##label CHE
      ##cross-references GB:U13978; NID:g555881; PID:g555882
GENETICS
   #introns      416/2
CLASSIFICATION   #superfamily vertebrate rhodopsin
KEYWORDS         neurotransmitter receptor; transmembrane protein
SUMMARY          #length 428  #molecular-weight 47398  #checksum 8085

  Query Match           65.4%;  Score 51;  DB 2;  Length 428;
  Best Local Similarity  40.0%;  Pred. No. 9.75e+00;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    122 YLAITAPLQY 131
          ||: :::|:|
Qy      3 YLSTSSSLHY 12


RESULT    6
ENTRY            S56830     #type complete
TITLE            probable purine nucleotide-binding protein YJL058c - yeast
                   (Saccharomyces cerevisiae)
ALTERNATE_NAMES  probable membrane protein YJL058c; protein J1141
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             05-May-1995 #sequence_revision 08-Sep-1995 #text_change
                   14-Nov-1997
ACCESSIONS       S56830
REFERENCE        S56793
   #authors      Pohl, T.M.; Aljinovic, G.
   #submission   submitted to the Protein Sequence Database, September 1995
   #accession    S56830
      ##molecule_type DNA
      ##residues      1-543 ##label TOV
      ##cross-references EMBL:Z49333; NID:g1008200; PID:g1008201; MIPS:YJL058c
GENETICS
   #map_position 10L
KEYWORDS         transmembrane protein
SUMMARY          #length 543  #molecular-weight 60840  #checksum 5268

  Query Match           65.4%;  Score 51;  DB 2;  Length 543;
  Best Local Similarity  63.6%;  Pred. No. 9.75e+00;
  Matches     7;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    182 ASFLTDSSTLH 192
          ||:|: ||:||
Qy      1 ASYLSTSSSLH 11


RESULT    7
ENTRY            S71629     #type complete
TITLE            sensory transduction histidine kinase dhkA - slime mold
                   (Dictyostelium discoideum)
ORGANISM         #formal_name Dictyostelium discoideum
DATE             29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change
                   24-Sep-1998
ACCESSIONS       S71629
REFERENCE        S71629
   #authors      Wang, N.; Shaulsky, G.; Escalante, R.; Loomis, W.F.
   #journal      EMBO J. (1996) 15:3890-3898
   #title        A two-component histidine kinase gene that functions in
                   Dictyostelium development.
   #cross-references MUID:96324397
   #accession    S71629
      ##status        nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-2150 ##label WAN
      ##cross-references EMBL:U42597
      ##experimental_source strain Ax4
GENETICS
   #gene         dhkA
   #map_position 6
CLASSIFICATION   #superfamily response regulator homology
KEYWORDS         autophosphorylation; phosphoprotein; phosphotransferase;
                   two-component regulatory system
FEATURE
   2027-2142          #domain response regulator homology #label RRH\
   2076               #binding_site phosphate (Asp) (covalent) #status
                        predicted
SUMMARY          #length 2150  #molecular-weight 239635  #checksum 5691

  Query Match           65.4%;  Score 51;  DB 2;  Length 2150;
  Best Local Similarity  70.0%;  Pred. No. 9.75e+00;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    180 NYLNNSSSLH 189
          :||: |||||
Qy      2 SYLSTSSSLH 11


RESULT    8
ENTRY            S46280     #type complete
TITLE            clpP protein - Chlamydomonas eugametos chloroplast
ORGANISM         #formal_name chloroplast Chlamydomonas eugametos
DATE             19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
                   01-Aug-1997
ACCESSIONS       S46280
REFERENCE        S46279
   #authors      Huang, C.; Wang, S.; Chen, L.; Lemieux, C.; Otis, C.; Turmel,
                   M.; Liu, X.Q.
   #journal      Mol. Gen. Genet. (1994) 244:151-159
   #title        The Chlamydomonas chloroplast clpP gene contains translated
                   large insertion sequences and is essential for cell growth.
   #cross-references MUID:94329067
   #accession    S46280
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-236 ##label HUA
GENETICS
   #genome       chloroplast
KEYWORDS         chloroplast
SUMMARY          #length 236  #molecular-weight 26638  #checksum 3940

  Query Match           64.1%;  Score 50;  DB 2;  Length 236;
  Best Local Similarity  40.0%;  Pred. No. 1.52e+01;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    177 YLTATETIHY 186
          ||::: ::||
Qy      3 YLSTSSSLHY 12


RESULT    9
ENTRY            A71648     #type complete
TITLE            pseudouridylate synthase I (truA) RP857 - Rickettsia
                   prowazekii
ORGANISM         #formal_name Rickettsia prowazekii
DATE             21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
                   21-Nov-1998
ACCESSIONS       A71648
REFERENCE        A71630
   #authors      Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
                   Sicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
                   Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
                   C.G.
   #journal      Nature (1998) 396:133-140
   #title        The genome sequence of Rickettsia prowazekii and the origin
                   of mitochondria.
   #accession    A71648
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
```

```
      ##molecule_type DNA
      ##residues       1-253 ##label AND
      ##cross-references GB:AJ235273; GB:AJ235269; NID:g3861237; PID:e1343126;
                        PID:g3861382
      ##experimental_source strain Madrid E
GENETICS
   #gene          truA; RP857
SUMMARY           #length 253  #molecular-weight 29075  #checksum 2010

  Query Match            64.1%;  Score 50;  DB 2;  Length 253;
  Best Local Similarity  60.0%;  Pred. No. 1.52e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    194 YISAPSFLHY 203
          |:|:::| |||
Qy      3 YLSTSSSLHY 12


RESULT   10
ENTRY             D65097     #type complete
TITLE             ygjR protein - Escherichia coli (strain K-12)
ORGANISM          #formal_name Escherichia coli
DATE              12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
                    13-Sep-1998
ACCESSIONS        D65097
REFERENCE         A64720
   #authors       Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                    Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                    Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                    Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                    Y.
   #journal       Science (1997) 277:1453-1462
   #title         The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession     D65097
      ##status         preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues       1-334 ##label BLAT
      ##cross-references GB:AE000390; GB:U00096; NID:g2367189; PID:g2367190;
                        UWGP:b3087
      ##experimental_source strain K-12, substrain MG1655
GENETICS
   #gene          ygjR
CLASSIFICATION    #superfamily Escherichia coli ygjR protein
SUMMARY           #length 334  #molecular-weight 36990  #checksum 2682

  Query Match            64.1%;  Score 50;  DB 2;  Length 334;
  Best Local Similarity  41.7%;  Pred. No. 1.52e+01;
  Matches     5;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db     72 AVYIASPNSLHF 83
          | |:::::|||:
Qy      1 ASYLSTSSSLHY 12


RESULT   11
ENTRY             S24320     #type complete
TITLE             Ig kappa chain precursor - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              02-Dec-1993 #sequence_revision 17-Nov-1995 #text_change
                    08-Sep-1997
ACCESSIONS        S24320
REFERENCE         S24319
   #authors       Aucouturier, P.; Khamlichi, A.A.; Preud'homme, J.L.; Bauwens,
                    M.; Touchard, G.; Cogne, M.
   #journal       Biochem. J. (1992) 285:149-152
   #title         Complementary DNA sequence of human amyloidogenic
                    immunoglobulin light-chain precursors.
   #cross-references MUID:92344562
   #accession     S24320
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-135 ##label AUC
      ##cross-references EMBL:X64133; NID:g32810; PID:g32811
      ##note           the authors translated the codon CAA for residue 122 as
                        Glu
CLASSIFICATION    #superfamily immunoglobulin V region; immunoglobulin homology
KEYWORDS          heterotetramer; immunoglobulin
SUMMARY           #length 135  #molecular-weight 14844  #checksum 5727

  Query Match            62.8%;  Score 49;  DB 2;  Length 135;
  Best Local Similarity  50.0%;  Pred. No. 2.36e+01;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db     73 ASYLETGVPLRF 84
          |||| |: :|::
Qy      1 ASYLSTSSSLHY 12


RESULT   12
ENTRY             A53533     #type complete
TITLE             phospholipid transfer protein precursor - human
ALTERNATE_NAMES   lipid transfer protein II; PLTP
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change
                    17-Mar-1999
ACCESSIONS        A53533
REFERENCE         A53533
   #authors       Day, J.R.; Albers, J.J.; Lofton-Day, C.E.; Gilbert, T.L.;
                    Ching, A.F.T.; Grant, F.J.; O'Hara, P.J.; Marcovina, S.M.;
                    Adolphson, J.L.
   #journal       J. Biol. Chem. (1994) 269:9388-9391
   #title         Complete cDNA encoding human phospholipid transfer protein
                    from human endothelial cells.
   #cross-references MUID:94179366
   #accession     A53533
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-493 ##label DAY
      ##cross-references GB:L26232; NID:g468325; PID:g468326
GENETICS
   #gene          GDB:PLTP
      ##cross-references GDB:340911; OMIM:172425
   #map_position 20pter-20qter
KEYWORDS          glycoprotein; phosphoprotein
FEATURE
   1-17                #domain signal sequence #status predicted #label SIG\
   18-493              #product phospholipid transfer protein #status
                         experimental #label MAT
SUMMARY           #length 493  #molecular-weight 54739  #checksum 6063

  Query Match            62.8%;  Score 49;  DB 2;  Length 493;
  Best Local Similarity  33.3%;  Pred. No. 2.36e+01;
  Matches     4;  Conservative    6;  Mismatches   2;  Indels   0;  Gaps   0;

Db    453 AGFLTIGADLHF 464
          |::|: :: ||:
Qy      1 ASYLSTSSSLHY 12


RESULT   13
ENTRY             F64713     #type complete
TITLE             protein-export membrane protein - Helicobacter pylori (strain
                    26695)
ORGANISM          #formal_name Helicobacter pylori
DATE              09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
                    12-Feb-1999
ACCESSIONS        F64713
REFERENCE         A64520
   #authors       Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
                    Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
                    H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
                    J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
                    Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
                    McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
```

```
                  Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
                  Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
                  Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
                  W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
                  C.M.; Venter, J.C.
   #journal       Nature (1997) 388:539-547
   #title         The complete genome sequence of the gastric pathogen
                    Helicobacter pylori.
   #cross-references MUID:97394467
   #accession     F64713
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-503 ##label TOM
      ##cross-references GB:AE000652; GB:AE000511; NID:g2314720; PID:g2314730;
                         TIGR:HP1550
GENETICS
   #start_codon  GTG
CLASSIFICATION   #superfamily protein export membrane protein secD
SUMMARY          #length 503  #molecular-weight 54247  #checksum 3320

  Query Match            62.8%;  Score 49;  DB 2;  Length 503;
  Best Local Similarity  60.0%;  Pred. No. 2.36e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db     38 YLSLASALEY 47
          ||| :|:| |
Qy      3 YLSTSSSLHY 12


RESULT   14
ENTRY            D71805     #type complete
TITLE            protein-export membrane protein - Helicobacter pylori (strain
                   J99)
ORGANISM         #formal_name Helicobacter pylori
   #variety      strain J99
DATE             12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                   05-Mar-1999
ACCESSIONS       D71805
REFERENCE        A71800
   #authors      Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                   Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                   B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                   Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                   Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                   G.F.; Trust, T.J.
   #journal      Nature (1999) 397:176-180
   #title        Genomic sequence comparison of two unrelated isolates of the
                   human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession    D71805
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-526 ##label ARN
      ##cross-references GB:AE001567; GB:AE001439; NID:g4156065; PID:g4156069
      ##experimental_source strain J99
GENETICS
   #gene         secD
CLASSIFICATION   #superfamily protein export membrane protein secD
SUMMARY          #length 526  #molecular-weight 56796  #checksum 5813

  Query Match            62.8%;  Score 49;  DB 2;  Length 526;
  Best Local Similarity  60.0%;  Pred. No. 2.36e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db     60 YLSLASALEY 69
          ||| :|:| |
Qy      3 YLSTSSSLHY 12


RESULT   15
ENTRY            T00353     #type fragment
TITLE            hypothetical protein KIAA0697 - human (fragment)
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
                   01-Feb-1999
ACCESSIONS       T00353
REFERENCE        Z14142
   #authors      Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka,
                   A.; Kotani, H.; Nomura, N.; Ohara, O.
   #journal      DNA Res. (1998) 5:169-176
   #title        Prediction of the coding sequences of unidentified human
                   genes. X. The complete sequences of 100 new cDNA clones
                   from brain which can code for large proteins in vitro.
   #accession    T00353
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-968 ##label ISH
      ##cross-references EMBL:AB014597; NID:d1204352; PID:d1032633
      ##experimental_source brain; clone HK04486
GENETICS
   #note         KIAA0697
SUMMARY          #length 968  #checksum 7201

  Query Match            62.8%;  Score 49;  DB 2;  Length 968;
  Best Local Similarity  54.5%;  Pred. No. 2.36e+01;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    653 SSYLPNSDPLH 663
          :|||: | :||
Qy      1 ASYLSTSSSLH 11


Search completed: Thu Sep  2 12:42:53 1999
Job time : 20 secs.
```

****

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:43:36 1999; MasPar time 4.40 Seconds
148.939 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-32
Description: (1-12) from US08599226.pep
Perfect Score: 78
Sequence: 1 ASYLSTSSSLHY 12

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.038; Variance 31.822; scale 0.755

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	59	75.6	989	11	Q63052	APOLIPOPROTEIN B (FRAG	4.42e-01
2	52	66.7	296	5	Q26364	HKB-SEGMENTATION GAP-G	1.08e+01
3	52	66.7	372	5	P91143	SIMILAR TO ACETYLTRANS	1.08e+01
4	52	66.7	531	8	O47582	NADH DEHYDROGENASE SUB	1.08e+01
5	52	66.7	1238	5	O61198	F15E6.6 PROTEIN.	1.08e+01
6	51	65.4	236	6	O77542	OLFACTORY RECEPTOR (FR	1.68e+01
7	51	65.4	238	6	O77543	OLFACTORY RECEPTOR (FR	1.68e+01
8	51	65.4	321	11	O70265	OLFACTORY RECEPTOR-LIK	1.68e+01
9	51	65.4	321	11	O70267	OLFACTORY RECEPTOR-LIK	1.68e+01
10	51	65.4	1494	11	O88902	PROTEIN TYROSINE PHOSP	1.68e+01
11	51	65.4	2150	5	Q23863	HISTIDINE KINASE A.	1.68e+01
12	51	65.4	2219	5	Q23388	ZK1067.2 PROTEIN.	1.68e+01
13	50	64.1	220	13	P70013	OLFACTORY RECEPTOR (FR	2.58e+01
14	50	64.1	357	5	O45973	Y6E2A.6 PROTEIN.	2.58e+01
15	50	64.1	419	10	O82449	P-HYDROXYPHENYLPYRUVAT	2.58e+01
16	50	64.1	444	5	P91141	SIMILAR TO ACETYLTRANS	2.58e+01
17	50	64.1	649	3	O60167	PROTEIN COMPLEX ASSEMB	2.58e+01
18	49	62.8	246	5	Q22895	SIMILAR TO FAMILY 1 OF	3.93e+01
19	49	62.8	503	2	O26074	PROTEIN-EXPORT MEMBRAN	3.93e+01
20	49	62.8	809	5	Q20702	SIMILAR TO MATRIN F/G.	3.93e+01
21	49	62.8	968	4	O75179	KIAA0697 PROTEIN (FRAG	3.93e+01
22	49	62.8	1392	10	O82493	T12H20.12 PROTEIN.	3.93e+01
23	48	61.5	176	9	Q37837	EXPRESSED IN UNINDUCED	5.97e+01
24	48	61.5	176	8	Q36294	CYTOCHROME B (FRAGMENT	5.97e+01
25	48	61.5	204	3	Q12097	HYPOTHETICAL 22.6 KD P	5.97e+01
26	48	61.5	301	5	Q22299	T07C5.4 PROTEIN.	5.97e+01
27	48	61.5	349	5	O17959	M01B2.5 PROTEIN.	5.97e+01
28	48	61.5	910	3	O59796	PUTATIVE RECEPTOR ASSO	5.97e+01
29	48	61.5	1046	2	P96156	CHITODEXTRINASE.	5.97e+01
30	48	61.5	1406	4	O15082	KIAA0377.	5.97e+01
31	47	60.3	98	6	O18812	DELTA 3 PIT-1 (FRAGMEN	9.00e+01
32	47	60.3	122	10	P93302	ORF122A.	9.00e+01
33	47	60.3	185	10	Q38688	LIGHT-HARVESTING CHLOR	9.00e+01
34	47	60.3	198	4	O75805	HOXA-9A.	9.00e+01
35	47	60.3	265	5	Q25144	HROX1.	9.00e+01
36	47	60.3	285	6	O02706	PIT-1, PARTIAL (FRAGME	9.00e+01
37	47	60.3	435	13	P70033	NUCLEAR ORPHAN RECEPTO	9.00e+01
38	47	60.3	463	5	Q22925	COSMID C50E3.	9.00e+01
39	47	60.3	613	3	O74954	HYPOTHETICAL 67.2 KD P	9.00e+01
40	47	60.3	646	14	Q65281	MAJOR CAPSID PROTEIN.	9.00e+01
41	47	60.3	725	2	Q52978	PHA[A,B,C,D,E,F,G] GEN	9.00e+01
42	47	60.3	932	5	O01623	SIMILAR TO LIGAND-GATE	9.00e+01
43	47	60.3	1101	5	Q22378	CODED FOR BY C. ELEGAN	9.00e+01
44	47	60.3	1746	6	Q29116	TENASCIN PRECURSOR (TN	9.00e+01
45	47	60.3	1943	5	O62341	R06F6.8B PROTEIN.	9.00e+01

ALIGNMENTS

```
RESULT    1
ID   Q63052      PRELIMINARY;      PRT;   989 AA.
AC   Q63052;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   APOLIPOPROTEIN B (FRAGMENT).
GN   APOB.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 89176719.
RA   REUBEN M.A., SVENSON K.L., DOOLITTLE M.H., JOHNSON D.F., LUSIS A.J.,
RA   ELOVSON J.;
RT   "Biosynthetic relationships between three rat apolipoprotein B
RT   peptides.";
RL   J. LIPID RES. 29:1337-1347(1988).
DR   EMBL; M27440; G623549; -.
KW   LIPOPROTEIN.
FT   NON_TER       1      1
SQ   SEQUENCE   989 AA;  113305 MW;  4EFC0E6E CRC32;

  Query Match           75.6%;  Score 59;  DB 11;  Length 989;
  Best Local Similarity  70.0%;  Pred. No. 4.42e-01;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    175 YLQASTSLHY 184
          || :|:||||
Qy      3 YLSTSSSLHY 12


RESULT    2
ID   Q26364      PRELIMINARY;      PRT;   296 AA.
AC   Q26364;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HKB=SEGMENTATION GAP-GENE HUCKEBEIN.
GN   HKB.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
```

```
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94268555.
RA   BRONNER G., CHU-LAGRAFF Q., DOE C.Q., COHEN B., WEIGEL D., TAUBERT H.,
RA   JACKLE H.;
RT   "Sp1/egr-like zinc-finger protein required for endoderm specification
RT   and germ-layer formation in Drosophila.";
RL   NATURE 369:664-668(1994).
DR   EMBL; S71230; G547124; -.
DR   FLYBASE; FBgn0001204; hkb.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 3.
DR   PFAM; PF00096; zf-C2H2; 3.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE   296 AA;  33598 MW;  5716DF75 CRC32;

  Query Match           66.7%;  Score 52;  DB 5;  Length 296;
  Best Local Similarity  50.0%;  Pred. No. 1.08e+01;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    127 APFLSAASDLYY 138
          |::||::| | |
Qy      1 ASYLSTSSSLHY 12


RESULT    3
ID   P91143      PRELIMINARY;      PRT;   372 AA.
AC   P91143;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO ACETYLTRANSFERASES.
GN   C37H5.2.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DAVIDSON S., GILLAM B.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U88315; G1825777; -.
DR   PFAM; PF00561; abhydrolase; 1.
KW   TRANSFERASE.
SQ   SEQUENCE   372 AA;  42139 MW;  5214F159 CRC32;

  Query Match           66.7%;  Score 52;  DB 5;  Length 372;
  Best Local Similarity  58.3%;  Pred. No. 1.08e+01;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    168 GGYLSTSYALKY 179
          ::|||||  :| |
Qy      1 ASYLSTSSSLHY 12


RESULT    4
ID   O47582      PRELIMINARY;      PRT;   531 AA.
AC   O47582;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   NADH DEHYDROGENASE SUBUNIT 5.
OS   ONCHOCERCA VOLVULUS.
OG   MITOCHONDRION.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; SPIRURIA; SPIRURIDA;
OC   FILARIOIDEA; ONCHOCERCIDAE; ONCHOCERCA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FOREST;
RA   KEDDIE E.M., UNNASCH T.R.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF015193; G2735946; -.
KW   MITOCHONDRION.
SQ   SEQUENCE   531 AA;  62891 MW;  47A734D9 CRC32;

  Query Match           66.7%;  Score 52;  DB 8;  Length 531;
  Best Local Similarity  50.0%;  Pred. No. 1.08e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    268 FLAIGSGLHY 277
          :|: :|:|||
Qy      3 YLSTSSSLHY 12


RESULT    5
ID   O61198      PRELIMINARY;      PRT;  1238 AA.
AC   O61198;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F15E6.6 PROTEIN.
GN   F15E6.6.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   MILLER N., STELLYES L., BRADSHAW H., KEPPLER D.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF038614; G2702437; -.
```

```
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW   IRON-SULFUR.
SQ   SEQUENCE   1238 AA;  135726 MW;  9408BB7C CRC32;

  Query Match           66.7%;  Score 52;  DB 5;  Length 1238;
  Best Local Similarity  70.0%;  Pred. No. 1.08e+01;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    702 YLETQSSLNY 711
          || | |||:|
Qy      3 YLSTSSSLHY 12


RESULT    6
ID   O77542      PRELIMINARY;      PRT;   236 AA.
AC   O77542;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   OLFACTORY RECEPTOR (FRAGMENT).
OS   SUS SCROFA (PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MATARAZZO V., TIRARD A., RENUCCI M., BELAICH A., CLEMENT J.L.;
RT   "Isolation of putative olfactory receptor sequences from pig nasal
RT   epithelium.";
RL   SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF042016; G3273633; -.
FT   NON_TER       1       1
FT   NON_TER     236     236
SQ   SEQUENCE   236 AA;  25564 MW;  B9651DA1 CRC32;

  Query Match           65.4%;  Score 51;  DB 6;  Length 236;
  Best Local Similarity  50.0%;  Pred. No. 1.68e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     66 YMAICSPLHY 75
          |::  |:|||
Qy      3 YLSTSSSLHY 12


RESULT    7
ID   O77543      PRELIMINARY;      PRT;   238 AA.
AC   O77543;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   OLFACTORY RECEPTOR (FRAGMENT).
OS   SUS SCROFA (PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MATARAZZO V., TIRARD A., RENUCCI M., BELAICH A., CLEMENT J.L.;
RT   "Isolation of putative olfactory receptor sequences from pig nasal
RT   epithelium.";
RL   SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF042017; G3273635; -.
FT   NON_TER       1       1
FT   NON_TER     238     238
SQ   SEQUENCE   238 AA;  25822 MW;  BEF1C7F3 CRC32;

  Query Match           65.4%;  Score 51;  DB 6;  Length 238;
  Best Local Similarity  50.0%;  Pred. No. 1.68e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     67 YMAICSPLHY 76
          |::  |:|||
Qy      3 YLSTSSSLHY 12


RESULT    8
ID   O70265      PRELIMINARY;      PRT;   321 AA.
AC   O70265;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   OLFACTORY RECEPTOR-LIKE PROTEIN.
GN   SCR D-8.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE; 98211958.
RA   WALENSKY L.D., RUAT M., BAKIN R.E., BLACKSHAW S., RONNETT G.V.,
RA   SNYDER S.H.;
RT   "Two novel odorant receptor families expressed in spermatids undergo
RT   5'-splicing.";
RL   J. BIOL. CHEM. 273:9378-9387(1998).
DR   EMBL; AF034897; G3153221; -.
SQ   SEQUENCE   321 AA;  35375 MW;  D85F971F CRC32;

  Query Match           65.4%;  Score 51;  DB 11;  Length 321;
  Best Local Similarity  50.0%;  Pred. No. 1.68e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    129 YMAICSPLHY 138
          |::  |:|||
Qy      3 YLSTSSSLHY 12


RESULT    9
ID   O70267      PRELIMINARY;      PRT;   321 AA.
AC   O70267;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   OLFACTORY RECEPTOR-LIKE PROTEIN.
GN   SCR D-9.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE; 98211958.
RA   WALENSKY L.D., RUAT M., BAKIN R.E., BLACKSHAW S., RONNETT G.V.,
RA   SNYDER S.H.;
RT   "Two novel odorant receptor families expressed in spermatids undergo
RT   5'-splicing.";
RL   J. BIOL. CHEM. 273:9378-9387(1998).
DR   EMBL; AF034899; G3153225; -.
SQ   SEQUENCE   321 AA;  35510 MW;  D060B840 CRC32;

  Query Match           65.4%;  Score 51;  DB 11;  Length 321;
  Best Local Similarity  50.0%;  Pred. No. 1.68e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    129 YMAICSPLHY 138
          |::  |:|||
Qy      3 YLSTSSSLHY 12


RESULT   10
ID   O88902      PRELIMINARY;      PRT;  1494 AA.
AC   O88902;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PROTEIN TYROSINE PHOSPHATASE TD14 (EC 3.1.3.48).
GN   PTP-TD14.
OS   RATTUS NORVEGICUS (RAT).
```

```
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98361981.
RA   CAO L., ZHANG L., RUIZ-LOZANO P., YANG Q., CHIEN K.R., GRAHAM R.M.,
RA   ZHOU M.;
RT   "A novel putative protein-tyrosine phosphatase contains a BRO1-like
RT   domain and suppresses Ha-ras-mediated transformation.";
RL   J. BIOL. CHEM. 273:21077-21083(1998).
DR   EMBL; AF077000; G3598974; -.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW   HYDROLASE.
SQ   SEQUENCE   1494 AA;  162931 MW;  9428F8BA CRC32;

  Query Match           65.4%;  Score 51;  DB 11;  Length 1494;
  Best Local Similarity  36.4%;  Pred. No. 1.68e+01;
  Matches    4;  Conservative    7;  Mismatches   0;  Indels   0;  Gaps   0;

Db    584 PFLGTAAPLHF 594
          ::|:|:::||:
Qy      2 SYLSTSSSLHY 12


RESULT   11
ID   Q23863      PRELIMINARY;      PRT;  2150 AA.
AC   Q23863;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HISTIDINE KINASE A.
GN   DHKA.
OS   DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC   EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE; 96324397.
RA   WANG N., SHAULSKY G., ESCALANTE R., LOOMIS W.F.;
RT   "A two-component histidine kinase gene that functions in
RT   Dictyostelium development.";
RL   EMBO J. 15:3890-3898(1996).
DR   EMBL; U42597; G1136289; -.
DR   PFAM; PF00072; response_reg; 1.
DR   PFAM; PF00512; signal; 1.
SQ   SEQUENCE   2150 AA;  239662 MW;  02F946CE CRC32;

  Query Match           65.4%;  Score 51;  DB 5;  Length 2150;
  Best Local Similarity  70.0%;  Pred. No. 1.68e+01;
  Matches    7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    180 NYLNNSSSLH 189
          :||: |||||
Qy      2 SYLSTSSSLH 11


RESULT   12
ID   Q23388      PRELIMINARY;      PRT;  2219 AA.
AC   Q23388;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   ZK1067.2 PROTEIN.
GN   ZK1067.2.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   THOMAS K.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z70038; E1351037; -.
SQ   SEQUENCE   2219 AA;  253649 MW;  59DE8B43 CRC32;

  Query Match           65.4%;  Score 51;  DB 5;  Length 2219;
  Best Local Similarity  50.0%;  Pred. No. 1.68e+01;
  Matches    6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    628 SSYLMRNGSLHF 639
          :|||  ::|||:
Qy      1 ASYLSTSSSLHY 12


RESULT   13
ID   P70013      PRELIMINARY;      PRT;  220 AA.
AC   P70013;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   OLFACTORY RECEPTOR (FRAGMENT).
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FREITAG J.;
RL   SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96112032.
RA   FREITAG J., KRIEGER J., STROTMAN J., BREER H.;
RT   "Two classes of olfactory receptors in Xenopus laevis.";
RL   NEURON 15:1383-1392(1995).
DR   EMBL; Y08351; E273932; -.
DR   PFAM; PF00001; 7tm_1; 1.
FT   NON_TER       1      1
FT   NON_TER     220    220
SQ   SEQUENCE   220 AA;  24765 MW;  EBD97044 CRC32;

  Query Match           64.1%;  Score 50;  DB 13;  Length 220;
  Best Local Similarity  50.0%;  Pred. No. 2.58e+01;
  Matches    5;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db     60 YLAICKPLHY 69
          ||:   :|||
Qy      3 YLSTSSSLHY 12


RESULT   14
ID   O45973      PRELIMINARY;      PRT;  357 AA.
AC   O45973;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   Y6E2A.6 PROTEIN.
GN   Y6E2A.6.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MATTHEWS L.;
RL   SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; AL021175; E1350810; -.
SQ   SEQUENCE   357 AA;  40700 MW;  9B5AEA9E CRC32;

  Query Match           64.1%;  Score 50;  DB 5;  Length 357;
  Best Local Similarity  50.0%;  Pred. No. 2.58e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    261 YLNGSAAFHY 270
          ||: |:::||
Qy      3 YLSTSSSLHY 12


RESULT   15
ID   O82449      PRELIMINARY;      PRT;   419 AA.
AC   O82449;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   P-HYDROXYPHENYLPYRUVATE DIOXYGENASE.
GN   PDS1.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE; 98369110.
RA   NORRIS S.R., SHEN X., DELLA PENNA D.;
RT   "Complementation of the arabidopsis pds1 mutation with the gene
RT   encoding p-hydroxyphenylpyruvate dioxygenase.";
RL   PLANT PHYSIOL. 117:1317-1323(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   DELLAPENNA D., NORRIS S.R., SHEN X.;
RL   SUBMITTED (APR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF060481; G3694811; -.
KW   DIOXYGENASE; PYRUVATE.
SQ   SEQUENCE   419 AA;  45854 MW;  60D99C92 CRC32;

  Query Match           64.1%;  Score 50;  DB 10;  Length 419;
  Best Local Similarity  58.3%;  Pred. No. 2.58e+01;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     86 ASYLLTSGDLRF 97
          |||| ||: |::
Qy      1 ASYLSTSSSLHY 12


Search completed: Thu Sep  2 12:44:04 1999
Job time : 28 secs.
```

```
**************************************************************************
 __  __  _____  _____  _____  _____  _____  _   _
|  \/  ||  __ \|  __ \|  __ \|  __ \|  __ \| | | |
| \  / || |__) | |__) | |__) | |__) | |__) | |_| |
| |\/| ||  ___/|  ___/|  ___/|  ___/|  ___/|  _  |
| |  | || |    | |    | |    | |    | |    | | | |
|_|  |_||_|    |_|    |_|    |_|    |_|    |_| |_|  (TM)
**************************************************************************
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:46:53 1999; MasPar time 3.13 Seconds
153.392 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-33
Description: (1-12) from US08599226.pep
Perfect Score: 74
Sequence: 1 ASFLSTSSSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.566; Variance 29.900; scale 0.822

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	54	73.0	469	2	D70048	ABC transporter (amin	2.14e+00
2	54	73.0	1069	2	T00040	BH-protocadherin PCDH	2.14e+00
3	54	73.0	1069	2	T00043	BH-protocadherin-a -	2.14e+00
4	54	73.0	1072	2	T00041	BH-protocadherin PCDH	2.14e+00
5	54	73.0	1200	2	T00042	BH-protocadherin PCDH	2.14e+00
6	53	71.6	112	4	S59333	hypothetical protein	3.44e+00
7	52	70.3	517	2	D37831	phenol 2-monooxygenas	5.50e+00
8	52	70.3	949	3	T03030	hypothetical protein	5.50e+00
9	50	67.6	296	2	S45336	finger protein, Spl/e	1.38e+01
10	50	67.6	311	2	S13808	protein-tyrosine kina	1.38e+01
11	50	67.6	346	2	S13809	protein-tyrosine kina	1.38e+01
12	50	67.6	346	2	S13807	protein-tyrosine kina	1.38e+01
13	50	67.6	464	2	B64970	hypothetical protein	1.38e+01
14	50	67.6	471	2	S01052	flavonol 3-O-glucosyl	1.38e+01
15	50	67.6	471	2	S08325	flavonol 3-O-glucosyl	1.38e+01
16	50	67.6	471	2	S01037	flavonol 3-O-glucosyl	1.38e+01
17	50	67.6	503	2	F64713	protein-export membra	1.38e+01
18	50	67.6	511	2	S47290	phenol 2-monooxygenas	1.38e+01
19	50	67.6	526	2	D71805	protein-export membra	1.38e+01
20	50	67.6	529	2	JC5533	scavenger receptor cl	1.38e+01
21	50	67.6	593	2	S49525	glycoprotein G - simi	1.38e+01
22	50	67.6	662	2	T01533	hypothetical protein	1.38e+01
23	50	67.6	1166	2	S06142	kinase-related transf	1.38e+01
24	49	66.2	155	2	B70169	hypothetical protein	2.16e+01
25	49	66.2	339	2	S73840	ribonucleotide reduct	2.16e+01
26	49	66.2	381	2	G71906	probable transcriptio	2.16e+01
27	49	66.2	439	1	KHQBTT	cysteine proteinase (	2.16e+01
28	49	66.2	509	1	A53920	scavenger receptor SR	2.16e+01
29	49	66.2	516	2	S44306	phenol hydroxylase -	2.16e+01
30	49	66.2	530	2	S62439	hypothetical protein	2.16e+01
31	48	64.9	500	2	S49302	AWJL218 protein - whe	3.37e+01
32	48	64.9	781	2	S43534	integrin beta3 - chic	3.37e+01
33	48	64.9	1495	2	S61023	hypothetical protein	3.37e+01
34	47	63.5	229	2	S33182	probable transport pr	5.20e+01
35	47	63.5	238	1	QQBE74	US4 protein - human h	5.20e+01
36	47	63.5	262	2	A53452	phenylalanine hydroxy	5.20e+01
37	47	63.5	265	2	S38380	Hroxl protein - Calif	5.20e+01
38	47	63.5	287	2	S57770	xyloglucan endo-1,4-b	5.20e+01
39	47	63.5	342	2	S63654	hypothetical protein	5.20e+01
40	47	63.5	457	2	E70081	purine-cytosine perme	5.20e+01
41	47	63.5	509	1	A48528	membrane glycoprotein	5.20e+01
42	47	63.5	568	1	A46339	hemagglutinin precurs	5.20e+01
43	47	63.5	668	2	H71312	probable ATP-dependen	5.20e+01
44	47	63.5	792	2	S03232	hypothetical protein	5.20e+01
45	47	63.5	828	2	S34695	hypothetical protein	5.20e+01

ALIGNMENTS

RESULT 1
ENTRY D70048 #type complete
TITLE ABC transporter (amino acid permease) homolog yvsH - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
ACCESSIONS D70048
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

```
   #cross-references MUID:98044033
   #accession    D70048
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-469 ##label KUN
      ##cross-references GB:Z99121; GB:AL009126; NID:g2635827; PID:e1186022;
                        PID:g2635847
      ##experimental_source strain 168
GENETICS
   #gene         yvsH
CLASSIFICATION   #superfamily L-lysine transport protein
SUMMARY          #length 469  #molecular-weight 50258  #checksum 4200

  Query Match           73.0%;  Score 54;  DB 2;  Length 469;
  Best Local Similarity  45.5%;  Pred. No. 2.14e+00;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    376 TFLTTAATLAY 386
          :||:|:::| |
Qy      2 SFLSTSSSLEY 12


RESULT    2
ENTRY            T00040     #type complete
TITLE            BH-protocadherin PCDH7 - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                   22-Jan-1999
ACCESSIONS       T00040
REFERENCE        Z14074
   #authors      Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
                   Sugano, S.
   #journal      Genomics (1998) 49:458-461
   #title        Cloning, expression analysis, and chromosomal localization of
                   BH-protocadherin (PCDH7), a novel member of the cadherin
                   superfamily.
   #accession    T00040
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-1069 ##label YOS
      ##cross-references EMBL:AB006755; NID:d1184677; PID:d1026122
      ##experimental_source clone BH-Pcdh-a
GENETICS
   #map_position 4p15
SUMMARY          #length 1069  #molecular-weight 116104  #checksum 9974

  Query Match           73.0%;  Score 54;  DB 2;  Length 1069;
  Best Local Similarity  60.0%;  Pred. No. 2.14e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    486 FLHTSTPLDY 495
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    3
ENTRY            T00043     #type complete
TITLE            BH-protocadherin-a - mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                   22-Jan-1999
ACCESSIONS       T00043
REFERENCE        Z14075
   #authors      Yoshida, K.
   #submission   submitted to the EMBL Data Library, August 1997
   #accession    T00043
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-1069 ##label YOS
      ##cross-references EMBL:AB006758; NID:d1227200; PID:d1033562
GENETICS
   #gene         pcdh7
   #map_position 5C3-D
SUMMARY          #length 1069  #molecular-weight 116313  #checksum 4821

  Query Match           73.0%;  Score 54;  DB 2;  Length 1069;
  Best Local Similarity  60.0%;  Pred. No. 2.14e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    486 FLHTSAPLDY 495
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    4
ENTRY            T00041     #type complete
TITLE            BH-protocadherin PCDH7 (clone BH-Pcdh-b) - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                   22-Jan-1999
ACCESSIONS       T00041
REFERENCE        Z14074
   #authors      Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
                   Sugano, S.
   #journal      Genomics (1998) 49:458-461
   #title        Cloning, expression analysis, and chromosomal localization of
                   BH-protocadherin (PCDH7), a novel member of the cadherin
                   superfamily.
   #accession    T00041
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-1072 ##label YOS
      ##cross-references EMBL:AB006756; NID:d1184678; PID:d1026123
      ##experimental_source clone BH-Pcdh-b
GENETICS
   #map_position 4p15
SUMMARY          #length 1072  #molecular-weight 116462  #checksum 9727

  Query Match           73.0%;  Score 54;  DB 2;  Length 1072;
  Best Local Similarity  60.0%;  Pred. No. 2.14e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    486 FLHTSTPLDY 495
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    5
ENTRY            T00042     #type complete
TITLE            BH-protocadherin PCDH7 (clone BH-Pcdh-c) - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                   22-Jan-1999
ACCESSIONS       T00042
REFERENCE        Z14074
   #authors      Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
                   Sugano, S.
   #journal      Genomics (1998) 49:458-461
   #title        Cloning, expression analysis, and chromosomal localization of
                   BH-protocadherin (PCDH7), a novel member of the cadherin
                   superfamily.
   #accession    T00042
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-1200 ##label YOS
      ##cross-references EMBL:AB006757; NID:d1184679; PID:d1026124
      ##experimental_source clone BH-Pcdh-c
GENETICS
   #map_position 4p15
SUMMARY          #length 1200  #molecular-weight 130337  #checksum 7152

  Query Match           73.0%;  Score 54;  DB 2;  Length 1200;
  Best Local Similarity  60.0%;  Pred. No. 2.14e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db    439 FLHTSTPLDY 448
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    6
ENTRY             S59333     #type complete
TITLE             hypothetical protein L3155 (L3149 internal orf) - yeast
                    (Saccharomyces cerevisiae)
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              29-Nov-1995 #sequence_revision 05-Sep-1996 #text_change
                    05-Sep-1996
ACCESSIONS        S59333
REFERENCE         S59313
   #authors       Delius, H.
   #submission    submitted to the EMBL Data Library, June 1995
   #description   36.8 kb of S.cerevisiae chromosome XII including ACE2, CKI1,
                    PDC5, SLS1, PUT1 and tRNA-Asp.
   #accession     S59333
      ##status         conceptual translation of pseudogene
      ##molecule_type DNA
      ##residues       1-112 ##label DEL
      ##cross-references EMBL:X91258; NID:g995686; PID:g995707
      ##experimental_source strain S288C
COMMENT     There is no evidence that this sequence is expressed.
GENETICS
   #map_position 12R
KEYWORDS          pseudogene
SUMMARY           #length 112  #molecular-weight 13520  #checksum 5962

  Query Match           71.6%;  Score 53;  DB 4;  Length 112;
  Best Local Similarity  70.0%;  Pred. No. 3.44e+00;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db     21 FLTISSSLAY 30
          ||: |||| |
Qy      3 FLSTSSSLEY 12


RESULT    7
ENTRY             D37831     #type complete
TITLE             phenol 2-monooxygenase (EC 1.14.13.7) chain P3 - Pseudomonas
                    sp. (strain CF600)
ORGANISM          #formal_name Pseudomonas sp.
DATE              14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change
                    31-Oct-1997
ACCESSIONS        D37831
REFERENCE         A37831
   #authors       Nordlund, I.; Powlowski, J.; Shingler, V.
   #journal       J. Bacteriol. (1990) 172:6826-6833
   #title         Complete nucleotide sequence and polypeptide analysis of
                    multicomponent phenol hydroxylase from Pseudomonas sp.
                    strain CF600.
   #cross-references MUID:91072230
   #accession     D37831
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-517 ##label NOR
      ##cross-references GB:M60276; GB:M37764; NID:g151449; PID:g151453
KEYWORDS          oxidoreductase
SUMMARY           #length 517  #molecular-weight 60522  #checksum 6342

  Query Match           70.3%;  Score 52;  DB 2;  Length 517;
  Best Local Similarity  60.0%;  Pred. No. 5.50e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    101 FLTAVSPLEY 110
          ||:: |:|||
Qy      3 FLSTSSSLEY 12


RESULT    8
ENTRY             T03030     #type fragment
TITLE             hypothetical protein KIAA0365 - human (fragment)
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              23-Mar-1999 #sequence_revision 23-Mar-1999 #text_change
                    23-Mar-1999
ACCESSIONS        T03030
REFERENCE         Z14651
   #authors       Lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson,
                    A.W.; Burkhart-Schultz, K.; Gordon, L.; Kyle, A.; Ramirez,
                    M.; Stilwagen, S.; Phan, H.; Velasco, N.; Garnes, J.;
                    Danganan, L.; Poundstone, P.; Christensen, M.; Georgescu,
                    A.; Avila, J.; Liu, S.; Attix, C.; Andreise, T.; Trankheim,
                    M.; Amico-Keller, G.; Coefield, J.; Duarte, S.; Lucas, S.;
                    Bruce, R.; Thomas, P.; Quan, G.; Kronmiller, B.; Arellano,
                    A.; Montgomery, M.; Ow, D.; Nolan, M.
   #submission    submitted to the EMBL Data Library, March 1998
   #description   Sequence analysis of an ~1 Mb region containing the MEF2B
                    gene in 19p12.
   #accession     T03030
      ##status         preliminary
      ##residues       1-949 ##label LAM
      ##cross-references EMBL:AC004447; NID:g2978446; PID:g2978447
SUMMARY           #length 949  #checksum 5364

  Query Match           70.3%;  Score 52;  DB 3;  Length 949;
  Best Local Similarity  70.0%;  Pred. No. 5.50e+00;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    475 FLSDENSLEY 484
          |||  :||||
Qy      3 FLSTSSSLEY 12


RESULT    9
ENTRY             S45336     #type complete
TITLE             finger protein, Sp1/egr-like - fruit fly (Drosophila sp.)
ORGANISM          #formal_name Drosophila sp.
DATE              19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
                    01-Aug-1997
ACCESSIONS        S45336
REFERENCE         S45336
   #authors       Broenner, G.; Chu-LaGraff, Q.; Doe, C.Q.; Cohen, B.; Weigel,
                    D.; Taubert, H.; Jaeckle, H.
   #journal       Nature (1994) 369:664-668
   #title         Sp1/egr-like zinc-finger protein required for endoderm
                    specification and germ-layer formation in Drosophila.
   #accession     S45336
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-296 ##label BRO
SUMMARY           #length 296  #molecular-weight 33598  #checksum 2521

  Query Match           67.6%;  Score 50;  DB 2;  Length 296;
  Best Local Similarity  58.3%;  Pred. No. 1.38e+01;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    127 APFLSAASDLYY 138
          |:|||::| | |
Qy      1 ASFLSTSSSLEY 12


RESULT   10
ENTRY             S13808     #type fragment
TITLE             protein-tyrosine kinase (EC 2.7.1.112) mrk (X chromosome) -
                    southern platyfish (fragment)
ALTERNATE_NAMES   melanoma-inducing protein
ORGANISM          #formal_name Xiphophorus maculatus #common_name southern
                    platyfish
DATE              21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
                    30-Jan-1998
ACCESSIONS        S13808
REFERENCE         S13807
   #authors       Adam, D.; Maeueler, W.; Schartl, M.
   #journal       Oncogene (1991) 6:73-80
```

```
   #title        Transcriptional activation of the melanoma inducing Xmrk
                   oncogene in Xiphophorus.
   #cross-references MUID:91125882
   #accession    S13808
      ##status        translation not shown
      ##molecule_type DNA
      ##residues      1-311 ##label ADA
      ##cross-references EMBL:X56318; NID:g65282; PID:g65283
GENETICS
   #map_position X
   #introns      52/3; 78/1; 127/1; 159/3; 205/3
CLASSIFICATION   #superfamily epidermal growth factor receptor; protein kinase
                   homology
KEYWORDS         ATP; phosphotransferase
FEATURE
   1-152              #domain protein kinase homology (fragment) #label KIN
SUMMARY          #length 311  #checksum 2822

  Query Match           67.6%;  Score 50;  DB 2;  Length 311;
  Best Local Similarity  50.0%;  Pred. No. 1.38e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    290 FLPAAENLEY 299
          ||::: :|||
Qy      3 FLSTSSSLEY 12


RESULT   11
ENTRY            S13809     #type fragment
TITLE            protein-tyrosine kinase (EC 2.7.1.112) mrk (Y chromosome) -
                   southern platyfish (fragment)
ALTERNATE_NAMES  melanoma-inducing protein
ORGANISM         #formal_name Xiphophorus maculatus #common_name southern
                   platyfish
DATE             21-Nov-1993 #sequence_revision 13-Mar-1997 #text_change
                   08-Sep-1997
ACCESSIONS       S13809
REFERENCE        S13807
   #authors      Adam, D.; Maeueler, W.; Schartl, M.
   #journal      Oncogene (1991) 6:73-80
   #title        Transcriptional activation of the melanoma inducing Xmrk
                   oncogene in Xiphophorus.
   #cross-references MUID:91125882
   #accession    S13809
      ##status        preliminary; translation not shown
      ##molecule_type DNA
      ##residues      1-346 ##label ADA
      ##cross-references EMBL:X56319; NID:g65284; PID:g65285
GENETICS
   #map_position Y
   #introns      52/3; 78/1; 127/1; 159/3; 205/3; 236/1
CLASSIFICATION   #superfamily epidermal growth factor receptor; protein kinase
                   homology
KEYWORDS         phosphotransferase
FEATURE
   1-152              #domain protein kinase homology (fragment) #label KIN
SUMMARY          #length 346  #checksum 797

  Query Match           67.6%;  Score 50;  DB 2;  Length 346;
  Best Local Similarity  50.0%;  Pred. No. 1.38e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    325 FLPAAENLEY 334
          ||::: :|||
Qy      3 FLSTSSSLEY 12


RESULT   12
ENTRY            S13807     #type fragment
TITLE            protein-tyrosine kinase (EC 2.7.1.112) mrk (invariably
                   present) - southern platyfish (fragment)
ALTERNATE_NAMES  melanoma-inducing protein
ORGANISM         #formal_name Xiphophorus maculatus #common_name southern
                   platyfish
DATE             21-Nov-1993 #sequence_revision 30-Jan-1998 #text_change
                   30-Jan-1998
ACCESSIONS       S13807
REFERENCE        S13807
   #authors      Adam, D.; Maeueler, W.; Schartl, M.
   #journal      Oncogene (1991) 6:73-80
   #title        Transcriptional activation of the melanoma inducing Xmrk
                   oncogene in Xiphophorus.
   #cross-references MUID:91125882
   #accession    S13807
      ##status        translation not shown
      ##molecule_type DNA
      ##residues      1-346 ##label ADA
      ##cross-references EMBL:X56317; NID:g65280; PID:g65281
GENETICS
   #gene         mrk
   #introns      52/3; 78/1; 127/1; 159/3; 205/3; 236/1
CLASSIFICATION   #superfamily epidermal growth factor receptor; protein kinase
                   homology
KEYWORDS         ATP; phosphotransferase
FEATURE
   1-152              #domain protein kinase homology (fragment) #label KIN
SUMMARY          #length 346  #checksum 1508

  Query Match           67.6%;  Score 50;  DB 2;  Length 346;
  Best Local Similarity  50.0%;  Pred. No. 1.38e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    325 FLPAAENLEY 334
          ||::: :|||
Qy      3 FLSTSSSLEY 12


RESULT   13
ENTRY            B64970     #type complete
TITLE            hypothetical protein b2043 - Escherichia coli (strain K-12)
ORGANISM         #formal_name Escherichia coli
DATE             12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
                   14-Nov-1997
ACCESSIONS       B64970
REFERENCE        A64720
   #authors      Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                   Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                   Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                   Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                   Y.
   #journal      Science (1997) 277:1453-1462
   #title        The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession    B64970
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-464 ##label BLAT
      ##cross-references GB:AE000295; GB:U00096; NID:g1788354; PID:g1788356;
                         UWGP:b2043
      ##experimental_source strain K-12, substrain MG1655
SUMMARY          #length 464  #molecular-weight 51315  #checksum 8372

  Query Match           67.6%;  Score 50;  DB 2;  Length 464;
  Best Local Similarity  45.5%;  Pred. No. 1.38e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     10 TFLTASSALAF 20
          :||::||:| :
Qy      2 SFLSTSSSLEY 12


RESULT   14
ENTRY            S01052     #type complete
TITLE            flavonol 3-O-glucosyltransferase (EC 2.4.1.91) (allele
                   Bz-McC) - maize
```

```
ALTERNATE_NAMES  UDPglucose flavonoid glucosyltransferase
ORGANISM         #formal_name Zea mays #common_name maize
DATE             31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
                   08-Sep-1997
ACCESSIONS       S01052; S08324
REFERENCE        S01037
   #authors      Ralston, E.J.; English, J.J.; Dooner, H.K.
   #journal      Genetics (1988) 119:185-197
   #title        Sequence of three bronze alleles of maize and correlation
                   with the genetic fine structure.
   #cross-references MUID:88284304
   #accession    S01052
      ##status        translation not shown
      ##molecule_type DNA
      ##residues      1-471 ##label RAL
      ##cross-references EMBL:X07940; NID:g22204; PID:g22205
REFERENCE        S08324
   #authors      Furtek, D.; Schiefelbein, J.W.; Johnston, F.; Nelson Jr.,
                   O.E.
   #journal      Plant Mol. Biol. (1988) 11:473-481
   #title        Sequence comparisons of three wild-type Bronze-1 alleles from
                   Zea mays.
   #accession    S08324
      ##status        translation not shown
      ##molecule_type DNA
      ##residues      1-471 ##label FUR
      ##cross-references EMBL:X13500; NID:g22364; PID:g1030071
GENETICS
   #gene         Bz1
   #map_position 9
   #introns      175/1
CLASSIFICATION   #superfamily flavonol O3-glucosyltransferase
KEYWORDS         glycosyltransferase; hexosyltransferase
SUMMARY          #length 471  #molecular-weight 48769  #checksum 6660

  Query Match           67.6%;  Score 50;  DB 2;  Length 471;
  Best Local Similarity  88.9%;  Pred. No. 1.38e+01;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      47 SFLSTASSL 55
           |||||:|||
Qy       2 SFLSTSSSL 10


RESULT   15
ENTRY            S08325     #type complete
TITLE            flavonol 3-O-glucosyltransferase (EC 2.4.1.91) (allele
                   BzMcC2) - maize
ALTERNATE_NAMES  UDPglucose flavonoid glucosyl-transferase
ORGANISM         #formal_name Zea mays #common_name maize
DATE             30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
                   08-Sep-1997
ACCESSIONS       S08325
REFERENCE        S08324
   #authors      Furtek, D.; Schiefelbein, J.W.; Johnston, F.; Nelson Jr.,
                   O.E.
   #journal      Plant Mol. Biol. (1988) 11:473-481
   #title        Sequence comparisons of three wild-type Bronze-1 alleles from
                   Zea mays.
   #accession    S08325
      ##status        translation not shown
      ##molecule_type DNA
      ##residues      1-471 ##label FUR
      ##cross-references EMBL:X13501; NID:g22361; PID:g295854
GENETICS
   #gene         Bz1
   #introns      175/1
CLASSIFICATION   #superfamily flavonol O3-glucosyltransferase
KEYWORDS         glycosyltransferase; hexosyltransferase
SUMMARY          #length 471  #molecular-weight 48621  #checksum 7439

  Query Match           67.6%;  Score 50;  DB 2;  Length 471;
  Best Local Similarity  88.9%;  Pred. No. 1.38e+01;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      47 SFLSTASSL 55
           |||||:|||
Qy       2 SFLSTSSSL 10


Search completed: Thu Sep  2 12:47:12 1999
Job time : 19 secs.
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

n on: Thu Sep 2 12:48:34 1999; MasPar time 1.45 Seconds
83.761 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-33
Description: (1-12) from US08599226.pep
Perfect Score: 74
Sequence: 1 ASFLSTSSSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 16.342; Variance 51.433; scale 0.318

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	52	70.3	1026	3	PCT-US95-0	Sequence 95, Applicati	3.63e+01
2	52	70.3	1026	1	US-07-998-	Sequence 95, Applicati	3.63e+01
3	52	70.3	1026	1	US-08-453-	Sequence 95, Applicati	3.63e+01
4	52	70.3	1026	1	US-08-453-	Sequence 95, Applicati	3.63e+01
5	52	70.3	1026	3	PCT-US93-1	Sequence 95, Applicati	3.63e+01
6	52	70.3	1026	2	US-08-268-	Sequence 95, Applicati	3.63e+01
7	52	70.3	1203	3	PCT-US95-0	Sequence 103, Applicat	3.63e+01
8	52	70.3	1203	1	US-07-998-	Sequence 103, Applicat	3.63e+01
9	52	70.3	1203	1	US-08-453-	Sequence 103, Applicat	3.63e+01
10	52	70.3	1203	1	US-08-453-	Sequence 103, Applicat	3.63e+01
11	52	70.3	1203	3	PCT-US93-1	Sequence 103, Applicat	3.63e+01
12	52	70.3	1203	2	US-08-268-	Sequence 103, Applicat	3.63e+01
13	50	67.6	23	2	US-08-985-	Sequence 11, Applicati	5.88e+01
14	50	67.6	445	2	US-08-985-	Sequence 2, Applicatio	5.88e+01
15	50	67.6	509	3	PCT-US95-0	Sequence 8, Applicatio	5.88e+01
16	49	66.2	509	3	PCT-US95-0	Sequence 4, Applicatio	7.47e+01
17	48	64.9	158	2	US-08-828-	Sequence 3, Applicatio	9.47e+01
18	48	64.9	159	2	US-08-828-	Sequence 1, Applicatio	9.47e+01
19	47	63.5	23	2	US-08-985-	Sequence 17, Applicati	1.20e+02
20	47	63.5	362	2	US-08-985-	Sequence 5, Applicatio	1.20e+02
21	46	62.2	218	3	PCT-US95-0	Sequence 4, Applicatio	1.52e+02
22	46	62.2	218	2	US-08-336-	Sequence 4, Applicatio	1.52e+02
23	46	62.2	238	2	US-08-928-	Sequence 5, Applicatio	1.52e+02
24	46	62.2	240	3	PCT-US95-0	Sequence 80, Applicati	1.52e+02
25	46	62.2	240	1	US-08-261-	Sequence 80, Applicati	1.52e+02
26	46	62.2	411	3	PCT-US95-0	Sequence 2, Applicatio	1.52e+02
27	46	62.2	411	2	US-08-336-	Sequence 2, Applicatio	1.52e+02
28	45	60.8	536	1	US-07-999-	Sequence 2, Applicatio	1.91e+02
29	45	60.8	536	1	US-08-426-	Sequence 2, Applicatio	1.91e+02
30	45	60.8	536	1	US-08-401-	Sequence 2, Applicatio	1.91e+02
31	45	60.8	536	1	US-08-354-	Sequence 2, Applicatio	1.91e+02
32	45	60.8	553	3	PCT-US94-0	Sequence 4, Applicatio	1.91e+02
33	45	60.8	554	2	US-08-464-	Sequence 1, Applicatio	1.91e+02
34	45	60.8	554	1	US-08-347-	Sequence 1, Applicatio	1.91e+02
35	45	60.8	554	3	PCT-US93-0	Sequence 2, Applicatio	1.91e+02
36	45	60.8	1041	1	US-08-220-	Sequence 4, Applicatio	1.91e+02
37	45	60.8	1041	1	US-08-413-	Sequence 4, Applicatio	1.91e+02
38	44	59.5	418	1	US-07-816-	Sequence 10, Applicati	2.41e+02
39	44	59.5	426	1	US-08-336-	Sequence 2, Applicatio	2.41e+02
40	44	59.5	426	3	PCT-US95-1	Sequence 2, Applicatio	2.41e+02
41	44	59.5	491	1	US-08-462-	Sequence 2, Applicatio	2.41e+02
42	44	59.5	1255	1	US-08-467-	Sequence 68, Applicati	2.41e+02
43	44	59.5	1255	2	US-08-468-	Sequence 68, Applicati	2.41e+02
44	44	59.5	1255	2	US-08-625-	Sequence 2, Applicatio	2.41e+02
45	44	59.5	1255	2	US-08-356-	Sequence 2, Applicatio	2.41e+02

ALIGNMENTS

```
RESULT    1
ID   PCT-US95-08071-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application PC/TUS9508071
XX
CC    Sequence 95, Application PC/TUS9508071
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  6300 Sears Tower, 233 S. Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/08071
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/12588
CC         FILING DATE:  23 DEC 1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/998,003
CC         FILING DATE:  29 DEC 1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Noland, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32149
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
```

```
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           70.3%;  Score 52;  DB 3;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT    2
ID   US-07-998-003A-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application US/07998003A
XX
CC    Sequence 95, Application US/07998003A
CC    Patent No. 5643781
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Bicknell
CC         STREET:  20 South Clark Street
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60603
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/998,003A
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5643781and, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  30903
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/346-5750
CC         TELEFAX:  312/984-9740
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           70.3%;  Score 52;  DB 1;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT    3
ID   US-08-453-695A-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application US/08453695A
XX
CC    Sequence 95, Application US/08453695A
CC    Patent No. 5708143
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  233 South Wacker, 6300 Sears Tower
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/453,695A
CC         FILING DATE:
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5708143and, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32658
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           70.3%;  Score 52;  DB 1;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT    4
ID   US-08-453-274B-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application US/08453274B
XX
CC    Sequence 95, Application US/08453274B
CC    Patent No. 5663300
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
```

```
CC         STREET:  6300 Sears Tower, 233 South Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  United States of America
CC         ZIP:  60606-6402
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/453,274B
CC         FILING DATE:  30-MAY-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5663300and, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32660
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           70.3%;  Score 52;  DB 1;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT    5
ID   PCT-US93-12588-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application PC/TUS9312588
XX
CC    Sequence 95, Application PC/TUS9312588
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  6300 Sears Tower, 233 S. Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/12588
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/998,003
CC         FILING DATE:  29 DEC 1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Noland, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  31811
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           70.3%;  Score 52;  DB 3;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT    6
ID   US-08-268-161A-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application US/08268161A
XX
CC    Sequence 95, Application US/08268161A
CC    Patent No. 5798224
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  233 South Wacker, 6300 Sears Tower
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/268,161A
CC         FILING DATE:  June 27, 1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Young J. Suh
CC         REGISTRATION NUMBER:  P-41,337
CC         REFERENCE/DOCKET NUMBER:  27866/32149
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           70.3%;  Score 52;  DB 2;  Length 1026;
```

```
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT    7
ID   PCT-US95-08071-103     STANDARD;     PRT;   1203 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 103, Application PC/TUS9508071
XX
CC    Sequence 103, Application PC/TUS9508071
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  6300 Sears Tower, 233 S. Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/08071
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/12588
CC         FILING DATE:  23 DEC 1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/998,003
CC         FILING DATE:  29 DEC 1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Noland, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32149
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  103:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1203 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1203 AA; 130280 MW; 7658453 CN;

  Query Match           70.3%;  Score 52;  DB 3;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT    8
ID   US-07-998-003A-103     STANDARD;     PRT;   1203 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 103, Application US/07998003A
XX
CC    Sequence 103, Application US/07998003A
CC    Patent No. 5643781
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Bicknell
CC         STREET:  20 South Clark Street
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60603
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/998,003A
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 564378land, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  30903
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/346-5750
CC         TELEFAX:  312/984-9740
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  103:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1203 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1203 AA; 130280 MW; 7658453 CN;

  Query Match           70.3%;  Score 52;  DB 1;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT    9
ID   US-08-453-695A-103     STANDARD;     PRT;   1203 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 103, Application US/08453695A
XX
CC    Sequence 103, Application US/08453695A
CC    Patent No. 5708143
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  233 South Wacker, 6300 Sears Tower
```

```
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/453,695A
CC         FILING DATE:
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5708143and, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32658
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  103:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1203 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1203 AA; 130280 MW; 7658453 CN;

  Query Match           70.3%;  Score 52;  DB 1;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT   10
ID   US-08-453-274B-103     STANDARD;     PRT;   1203 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 103, Application US/08453274B
XX
CC    Sequence 103, Application US/08453274B
CC    Patent No. 5663300
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
CC         STREET:  6300 Sears Tower, 233 South Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  United States of America
CC         ZIP:  60606-6402
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/453,274B
CC         FILING DATE:  30-MAY-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5663300and, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32660
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  103:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1203 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1203 AA; 130280 MW; 7658453 CN;

  Query Match           70.3%;  Score 52;  DB 1;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT   11
ID   PCT-US93-12588-103     STANDARD;     PRT;   1203 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 103, Application PC/TUS9312588
XX
CC    Sequence 103, Application PC/TUS9312588
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  6300 Sears Tower, 233 S. Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/12588
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/998,003
CC         FILING DATE:  29 DEC 1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Noland, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  31811
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  103:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1203 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1203 AA; 130280 MW; 7658453 CN;

  Query Match           70.3%;  Score 52;  DB 3;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT   12
ID   US-08-268-161A-103     STANDARD;     PRT;   1203 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 103, Application US/08268161A
XX
CC    Sequence 103, Application US/08268161A
CC    Patent No. 5798224
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  233 South Wacker, 6300 Sears Tower
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/268,161A
CC         FILING DATE:  June 27, 1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Young J. Suh
CC         REGISTRATION NUMBER:  P-41,337
CC         REFERENCE/DOCKET NUMBER:  27866/32149
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  103:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1203 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1203 AA; 130280 MW; 7658453 CN;

  Query Match           70.3%;  Score 52;  DB 2;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 3.63e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT   13
ID   US-08-985-090-11     STANDARD;     PRT;   23 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 11, Application US/08985090
XX
CC    Sequence 11, Application US/08985090
CC    Patent No. 5885893
CC     GENERAL INFORMATION:
CC       APPLICANT:  Andrew D.J. Goodearl
CC       TITLE OF INVENTION:  MUSCARINIC RECEPTORS AND USES THEREFOR
CC       NUMBER OF SEQUENCES:  28
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  LAHIVE & COCKFIELD, LLP
CC         STREET:  28 State Street
CC         CITY:  Boston
CC         STATE:  Massachusetts
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/985,090
CC         FILING DATE:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Jean M. Silveri
CC         REGISTRATION NUMBER:  39,030
CC         REFERENCE/DOCKET NUMBER:  MNI-032
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617)227-7400
CC         TELEFAX:  (617)742-4214
CC     INFORMATION FOR SEQ ID NO:  11:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  23 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   23 AA; 2719 MW; 3615 CN;

  Query Match           67.6%;  Score 50;  DB 2;  Length 23;
  Best Local Similarity  60.0%;  Pred. No. 5.88e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db      2 FLITASTLEF 11
          || |:|:||:
Qy      3 FLSTSSSLEY 12


RESULT   14
ID   US-08-985-090-2     STANDARD;     PRT;   445 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08985090
XX
CC    Sequence 2, Application US/08985090
CC    Patent No. 5885893
CC     GENERAL INFORMATION:
CC       APPLICANT:  Andrew D.J. Goodearl
CC       TITLE OF INVENTION:  MUSCARINIC RECEPTORS AND USES THEREFOR
CC       NUMBER OF SEQUENCES:  28
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  LAHIVE & COCKFIELD, LLP
CC         STREET:  28 State Street
CC         CITY:  Boston
CC         STATE:  Massachusetts
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/985,090
CC         FILING DATE:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Jean M. Silveri
CC         REGISTRATION NUMBER:  39,030
CC         REFERENCE/DOCKET NUMBER:  MNI-032
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617)227-7400
CC         TELEFAX:  (617)742-4214
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  445 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   445 AA; 48671 MW; 1062278 CN;

  Query Match           67.6%;  Score 50;  DB 2;  Length 445;
  Best Local Similarity  60.0%;  Pred. No. 5.88e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    198 FLITASTLEF 207
          || |:|:||:
Qy      3 FLSTSSSLEY 12


RESULT   15
ID   PCT-US95-07721-8     STANDARD;     PRT;   509 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application PC/TUS9507721
XX
CC    Sequence 8, Application PC/TUS9507721
CC     GENERAL INFORMATION:
CC       APPLICANT:  Massachusetts Institute of Technology
CC       TITLE OF INVENTION:  Class BI and CI Scavenger Receptors
CC       NUMBER OF SEQUENCES:  8
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Patrea L. Pabst
CC         STREET:  2800 One Atlantic Center
CC         STREET:  1201 West Peachtree Street
CC         CITY:  Atlanta
CC         STATE:  Georgia
CC         COUNTRY:  USA
CC         ZIP:  30309-3450
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/07721
CC         FILING DATE:
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Pabst, Patrea L.
CC         REGISTRATION NUMBER:  31,284
CC         REFERENCE/DOCKET NUMBER:  MIT6620
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (404) 873-8794
CC         TELEFAX:  (404) 873-8795
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  509 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       FEATURE:
CC         NAME/KEY:  misc. feature
CC         LOCATION:  1..509
CC         OTHER INFORMATION:  /Function = "Amino acid sequence for
CC         OTHER INFORMATION:  the murine Scavenger Receptor Class BI."
SQ   SEQUENCE   509 AA; 56754 MW; 1451442 CN;

  Query Match           67.6%;  Score 50;  DB 3;  Length 509;
  Best Local Similarity  50.0%;  Pred. No. 5.88e+01;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    263 APFMTPESSLEF 274
          |:|::  ||||:
Qy      1 ASFLSTSSSLEY 12


Search completed: Thu Sep  2 12:48:42 1999
Job time : 8 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:47:55 1999; MasPar time 4.49 Seconds
145.743 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-33
Description: (1-12) from US08599226.pep
Perfect Score: 74
Sequence: 1 ASFLSTSSSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.159; Variance 28.834; scale 0.838

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	55	74.3	475	13	O93514	AXIAL PROTOCADHERIN (F	1.19e+00
2	54	73.0	469	2	O32204	YVSH PROTEIN.	1.96e+00
3	54	73.0	1035	13	O57537	NF-PROTOCADHERIN.	1.96e+00
4	54	73.0	1069	4	O60245	PCDH7 (BH-PCDH)A.	1.96e+00
5	54	73.0	1069	11	O88185	BH-PROTOCADHERIN-A.	1.96e+00
6	54	73.0	1072	4	O60246	PCDH7 (BH-PCDH)B.	1.96e+00
7	54	73.0	1200	4	O60247	PCDH7 (BH-PCDH)C.	1.96e+00
8	53	71.6	112	3	Q07255	INTERNAL ORF OF L3149	3.20e+00
9	52	70.3	809	5	Q20702	SIMILAR TO MATRIN F/G.	5.19e+00
10	52	70.3	939	4	O15071	KIAA0365 (FRAGMENT).	5.19e+00
11	52	70.3	949	4	O60369	KIAA0365 (FRAGMENT).	5.19e+00
12	52	70.3	1026	4	Q08174	PROTOCADHERIN 42 PRECU	5.19e+00
13	51	68.9	649	3	O60167	PROTEIN COMPLEX ASSEMB	8.37e+00
14	50	67.6	60	14	Q69367	GLYCOPROTEIN GG (FRAGM	1.34e+01
15	50	67.6	198	4	O75805	HOXA-9A.	1.34e+01
16	50	67.6	296	5	Q26364	HKB=SEGMENTATION GAP-G	1.34e+01
17	50	67.6	311	13	Q99162	MELANOMA RECEPTOR PROT	1.34e+01
18	50	67.6	346	13	P11776	PROTO-ONCOGENE RECEPTO	1.34e+01
19	50	67.6	471	5	Q26199	SERA-3 (FRAGMENT).	1.34e+01
20	50	67.6	503	2	O26074	PROTEIN-EXPORT MEMBRAN	1.34e+01
21	50	67.6	509	11	Q61009	SCAVENGER RECEPTOR CLA	1.34e+01
22	50	67.6	509	11	P97943	SCAVENGER RECEPTOR CLA	1.34e+01
23	50	67.6	509	6	O18824	SCAVENGER RECEPTOR CLA	1.34e+01
24	50	67.6	509	11	O88548	TYPE II PNEUMOCYTE CD3	1.34e+01
25	50	67.6	511	2	Q43981	PHENOLHYDROXYLASE COMP	1.34e+01
26	50	67.6	511	2	O32431	DMS OXYGENASE COMPONEN	1.34e+01
27	50	67.6	593	14	Q87093	GLYCOPROTEIN G (HOMOLO	1.34e+01
28	50	67.6	599	14	Q80929	REPLICATION PROTEIN E1	1.34e+01
29	50	67.6	662	10	O23064	BAC IG005I10.	1.34e+01
30	49	66.2	155	2	O51505	HYPOTHETICAL 18.1 KD P	2.13e+01
31	49	66.2	303	14	Q84594	SIMILAR TO PBCV-1 SERI	2.13e+01
32	49	66.2	442	2	O50151	UVRA PROTEIN.	2.13e+01
33	49	66.2	509	11	Q60417	CD36-RELATED CLASS B S	2.13e+01
34	49	66.2	516	2	Q52173	PHENOL HYDROXYLASE P3	2.13e+01
35	49	66.2	643	3	O42633	FATTY ACID TRANSPORTER	2.13e+01
36	49	66.2	664	10	O22716	F8A5.30 PROTEIN.	2.13e+01
37	49	66.2	890	5	Q20388	SIMILAR TO LIN-15B PRO	2.13e+01
38	49	66.2	1476	13	Q90285	PUTATIVE EXTRACELLULAR	2.13e+01
39	48	64.9	224	2	O08247	HSP18 TRANSCRIPTION RE	3.37e+01
40	48	64.9	349	5	O17959	M01B2.5 PROTEIN.	3.37e+01
41	48	64.9	376	2	O85449	HYPOTHETICAL 42.6 KD P	3.37e+01
42	48	64.9	500	10	Q41537	(SUBCLONE PAWJL218) AW	3.37e+01
43	48	64.9	781	13	Q92071	INTEGRIN BETA3.	3.37e+01
44	48	64.9	932	5	O01623	SIMILAR TO LIGAND-GATE	3.37e+01
45	48	64.9	1495	3	Q12280	CHROMOSOME XVI READING	3.37e+01

ALIGNMENTS

RESULT 1
ID O93514 PRELIMINARY; PRT; 475 AA.
AC O93514;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE AXIAL PROTOCADHERIN (FRAGMENT).
GN AXPC.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RA YAMAMOTO A., DEROBERTIS E.M.;
RT "Xenopus axial protocadherin.";
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AF053469; G3598688; -.
FT NON_TER 1 1
FT NON_TER 475 475
SQ SEQUENCE 475 AA; 52268 MW; 2A681544 CRC32;

Query Match 74.3%; Score 55; DB 13; Length 475;
Best Local Similarity 60.0%; Pred. No. 1.19e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Db    341 FLQTTTSLDY 350
          || |::||:|
Qy      3 FLSTSSSLEY 12
```

RESULT 2
ID O32204 PRELIMINARY; PRT; 469 AA.
AC O32204;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE YVSH PROTEIN.
GN YVSH.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;

```
RX   MEDLINE; 98044033.
RA   KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA   AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA   BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA   BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA   CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA   DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA   ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA   FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA   GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA   GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA   HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA   JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA   KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA   KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA   LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA   MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA   NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA   PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA   PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA   RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA   SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA   SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
RA   SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA   TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA   TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA   VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA   WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA   YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   NATURE 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   WIPAT A., BRIGNELL C.S., GUY J.B., ROSE M., EMMERSON P.T.,
RA   HARWOOD C.R.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z99121; E1186022; -.
DR   EMBL; AJ223978; E1249790; -.
DR   PFAM; PF00324; aa_permeases; 1.
SQ   SEQUENCE   469 AA;  50258 MW;  49186162 CRC32;

  Query Match           73.0%;  Score 54;  DB 2;  Length 469;
  Best Local Similarity  45.5%;  Pred. No. 1.96e+00;
  Matches    5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    376 TFLTTAATLAY 386
          :||:|:::| |
Qy      2 SFLSTSSSLEY 12


RESULT    3
ID   O57537      PRELIMINARY;      PRT;  1035 AA.
AC   O57537;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE   NF-PROTOCADHERIN.
GN   NFPC.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BRADLEY R.S., ESPESETH A., KINTNER C.;
RL   CURR. BIOL. 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF043643; G2852363; -.
DR   PROSITE; PS00232; CADHERIN; 6.
KW   CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ   SEQUENCE   1035 AA;  113713 MW;  7E4D3C4E CRC32;

  Query Match           73.0%;  Score 54;  DB 13;  Length 1035;
  Best Local Similarity  60.0%;  Pred. No. 1.96e+00;
  Matches    6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    458 FLHTSAPLDY 467
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    4
ID   O60245      PRELIMINARY;      PRT;  1069 AA.
AC   O60245;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   PCDH7 (BH-PCDH)A.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   GENOMICS 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006755; D1026122; -.
DR   PROSITE; PS00232; CADHERIN; 6.
KW   CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ   SEQUENCE   1069 AA;  116104 MW;  F1732B30 CRC32;

  Query Match           73.0%;  Score 54;  DB 4;  Length 1069;
  Best Local Similarity  60.0%;  Pred. No. 1.96e+00;
  Matches    6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    486 FLHTSTPLDY 495
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    5
ID   O88185      PRELIMINARY;      PRT;  1069 AA.
AC   O88185;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   BH-PROTOCADHERIN-A.
GN   PCDH7.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98277460.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RT   "Cloning, expression analysis, and chromosomal localization of
RT   BH-protocadherin (PCDH7), a novel member of the cadherin
RT   superfamily.";
RL   GENOMICS 49:458-461(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., HIDA M., WATANABE M., YAMAGUCHI R., TATEYAMA S.,
RA   SUGANO S.;
RT   "cDNA cloning and chromosomal mapping of mouse BH-protocadherin.";
RL   DNA SEQ. 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006758; D1033562; -.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ   SEQUENCE   1069 AA;  116314 MW;  0F3F60C6 CRC32;
```

```
  Query Match           73.0%;  Score 54;  DB 11;  Length 1069;
  Best Local Similarity  60.0%;  Pred. No. 1.96e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    486 FLHTSAPLDY 495
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    6
ID   O60246      PRELIMINARY;      PRT;  1072 AA.
AC   O60246;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   PCDH7 (BH-PCDH)B.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
     CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   GENOMICS 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006756; D1026123; -.
DR   PROSITE; PS00232; CADHERIN; 6.
KW   CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ   SEQUENCE   1072 AA;  116463 MW;  A3DF367C CRC32;

  Query Match           73.0%;  Score 54;  DB 4;  Length 1072;
  Best Local Similarity  60.0%;  Pred. No. 1.96e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    486 FLHTSTPLDY 495
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    7
ID   O60247      PRELIMINARY;      PRT;  1200 AA.
AC   O60247;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   PCDH7 (BH-PCDH)C.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
     CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   GENOMICS 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006757; D1026124; -.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ   SEQUENCE   1200 AA;  130337 MW;  56F1CD33 CRC32;

  Query Match           73.0%;  Score 54;  DB 4;  Length 1200;
  Best Local Similarity  60.0%;  Pred. No. 1.96e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    439 FLHTSTPLDY 448
          || ||::|:|
Qy      3 FLSTSSSLEY 12


RESULT    8
ID   Q07255      PRELIMINARY;      PRT;  112 AA.
AC   Q07255;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   INTERNAL ORF OF L3149 PROBABLY NOT CODING.
GN   L3155.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RA   DELIUS H.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; X91258; G995707; -.
SQ   SEQUENCE   112 AA;  13520 MW;  02B2E1AC CRC32;

  Query Match           71.6%;  Score 53;  DB 3;  Length 112;
  Best Local Similarity  70.0%;  Pred. No. 3.20e+00;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db     21 FLTISSSLAY 30
          ||: |||| |
Qy      3 FLSTSSSLEY 12


RESULT    9
ID   Q20702      PRELIMINARY;      PRT;  809 AA.
AC   Q20702;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO MATRIN F/G.
GN   F53B1.8.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   MILLER N.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U40953; G1072256; -.
SQ   SEQUENCE   809 AA;  89741 MW;  A01E9234 CRC32;

  Query Match           70.3%;  Score 52;  DB 5;  Length 809;
  Best Local Similarity  54.5%;  Pred. No. 5.19e+00;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db    297 GFLGASVSLQY 307
          :||::| ||:|
Qy      2 SFLSTSSSLEY 12


RESULT   10
ID   O15071      PRELIMINARY;      PRT;   939 AA.
AC   O15071;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   KIAA0365 (FRAGMENT).
GN   KIAA0365.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97349984.
RA   NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA   TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA RES. 4:141-150(1997).
DR   EMBL; AB002363; D1021661; -.
FT   NON_TER       1      1
SQ   SEQUENCE   939 AA;  103677 MW;  4B75FDB9 CRC32;

  Query Match           70.3%;  Score 52;  DB 4;  Length 939;
  Best Local Similarity  70.0%;  Pred. No. 5.19e+00;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    475 FLSDENSLEY 484
          |||  :||||
Qy      3 FLSTSSSLEY 12


RESULT   11
ID   O60369      PRELIMINARY;      PRT;   949 AA.
AC   O60369;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   KIAA0365 (FRAGMENT).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LAMERDIN J.E., MCCREADY P.M., SKOWRONSKI E., ADAMSON A.W.,
RA   BURKHART-SCHULTZ K., GORDON L., KYLE A., RAMIREZ M., STILWAGEN S.,
RA   PHAN H., VELASCO N., GARNES J., DANGANAN L., POUNDSTONE P.,
RA   CHRISTENSEN M., GEORGESCU A., AVILA J., LIU S., ATTIX C., ANDREISE T.,
RA   TRANKHEIM M., AMICO-KELLER G., COEFIELD J., DUARTE S., LUCAS S.,
RA   BRUCE R., THOMAS P., QUAN G., KRONMILLER B., ARELLANO A.,
RA   MONTGOMERY M., OW D., NOLAN M., TRONG S., KOBAYASHI A., OLSEN A.O.,
RA   CARRANO A.V.;
RL   SUBMITTED (MAR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AC004447; G2978447; -.
FT   NON_TER       1      1
SQ   SEQUENCE   949 AA;  105111 MW;  1C0F3FB3 CRC32;

  Query Match           70.3%;  Score 52;  DB 4;  Length 949;
  Best Local Similarity  70.0%;  Pred. No. 5.19e+00;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    475 FLSDENSLEY 484
          |||  :||||
Qy      3 FLSTSSSLEY 12


RESULT   12
ID   Q08174      PRELIMINARY;      PRT;  1026 AA.
AC   Q08174;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PROTOCADHERIN 42 PRECURSOR (PC42) (CADHERIN-LIKE PROTEIN).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 93285094.
RA   SANO K., TANIHARA H., HEIMARK R.L., OBATA S., DAVIDSON M., ST JOHN T.,
RA   TAKETANI S., SUZUKI S.;
RT   "Protocadherins: a large family of cadherin-related molecules in
RT   central nervous system.";
RL   EMBO J. 12:2249-2256(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTION PROCESSES AND
CC       IN CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC       FOUND AT CELL-CELL BOUNDARIES AND PROBABLY AT CELL-MATRIX
CC       BOUNDARIES.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE BRAIN AND NEURO-GLIAL
CC       CELLS.
CC   -!- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION IN ADULTS.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY OF CELL ADHESION
CC       MOLECULES. CONTAINS 7 CADHERINS-TYPE REPEATS.
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF THIS PROTEIN THAT DIFFER
CC       AT THE C-TERMINUS ARE PRODUCED BY ALTERNATIVE SPLICING.
DR   EMBL; L11370; G387675; -.
DR   EMBL; L11369; G387673; -.
DR   PFAM; PF00028; cadherin; 7.
KW   CELL ADHESION; TRANSMEMBRANE; CALCIUM-BINDING; ALTERNATIVE SPLICING;
KW   REPEAT; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?   1026       PROTOCADHERIN 42.
FT   DOMAIN       42    818       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    819    839       POTENTIAL.
FT   DOMAIN      840   1026       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       42    146       CADHERIN 1.
FT   REPEAT      147    246       CADHERIN 2.
FT   REPEAT      247    353       CADHERIN 3.
FT   REPEAT      354    472       CADHERIN 4.
FT   REPEAT      473    578       CADHERIN 5.
FT   REPEAT      579    681       CADHERIN 6.
FT   REPEAT      682    818       CADHERIN 7.
FT   VARSPLIC   1000   1026       VGQPFQLSTPQPLPHPYHGAIWTEVWE ->
FT                                LPHRRVTFSA TSQAQELQDPSQHSYYDSGLE
FT                                (IN PC42-8).
SQ   SEQUENCE   1026 AA;  111270 MW;  84A6E132 CRC32;

  Query Match           70.3%;  Score 52;  DB 4;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 5.19e+00;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT   13
ID   O60167      PRELIMINARY;      PRT;   649 AA.
AC   O60167;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   PROTEIN COMPLEX ASSEMBLY PROTEIN.
GN   SPBC19F8.03C.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
```

```
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   BECK A., REINHARDT R., WOOD V., RAJANDREAM M.A., BARRELL B.G.;
RL   SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AL023594; E1293401; -.
SQ   SEQUENCE   649 AA;  72985 MW;  9C207DB2 CRC32;

  Query Match           68.9%;  Score 51;  DB 3;  Length 649;
  Best Local Similarity  54.5%;  Pred. No. 8.37e+00;
  Matches    6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    234 NYLSTARSLEF 244
          ::|||: |||:
Qy      2 SFLSTSSSLEY 12


RESULT   14
    Q69367      PRELIMINARY;      PRT;    60 AA.
    Q69367;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   GLYCOPROTEIN GG (FRAGMENT).
OS   SIMIAN HERPESVIRUS SA8.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC   ALPHAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93298054.
RA   EBERLE R., ZHANG M., BLACK D.;
RT   "Gene mapping and sequence analysis of the unique short region of the
RT   simian herpesvirus SA 8 genome.";
RL   ARCH. VIROL. 130:391-411(1993).
DR   EMBL; L05608; G331077; -.
FT   NON_TER       1      1
SQ   SEQUENCE   60 AA;  6623 MW;  5CE36ED8 CRC32;

  Query Match           67.6%;  Score 50;  DB 14;  Length 60;
  Best Local Similarity  36.4%;  Pred. No. 1.34e+01;
  Matches    4;  Conservative    7;  Mismatches   0;  Indels   0;  Gaps   0;

Db      4 PFLTASPGLDF 14
          :||::|::|::
Qy      2 SFLSTSSSLEY 12


RESULT   15
ID   O75805      PRELIMINARY;      PRT;  198 AA.
AC   O75805;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HOXA-9A.
GN   HOXA-9.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98290548.
RA   KIM M.H., CHANG H.H., SHIN C., CHO M., PARK D., PARK H.W.;
RT   "Genomic structure and sequence analysis of human HOXA-9.";
RL   DNA CELL BIOL. 17:407-414(1998).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; U81511; G3237298; -.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
KW   HOMEOBOX; NUCLEAR PROTEIN; DNA-BINDING.
SQ   SEQUENCE   198 AA;  22874 MW;  307982CA CRC32;

  Query Match           67.6%;  Score 50;  DB 4;  Length 198;
  Best Local Similarity  41.7%;  Pred. No. 1.34e+01;
  Matches    5;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db     42 STFLSTGTTLPF 53
          ::||||:::| :
Qy      1 ASFLSTSSSLEY 12


Search completed: Thu Sep  2 12:48:16 1999
Job time : 21 secs.
```

```
OC   ALPHAHERPESVIRINAE; SIMPLEXVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85160822.
RA   MCGEOCH D.J., DOLAN A., DONALD S., RIXON F.J.;
RT   "Sequence determination and genetic content of the short unique
RT   region in the genome of herpes simplex virus type 1.";
RL   J. MOL. BIOL. 181:1-13(1985).
CC   -!- THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND 2: GH, GB, GC,
CC       GG, GD, GI, AND GE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00036; G291500; -.
DR   EMBL; X14112; G59562; -.
DR   EMBL; X14112; E312376; -.
DR   EMBL; X02138; G59870; -.
DR   PIR; A05239; QQBE74.
KW   GLYCOPROTEIN.
FT   CARBOHYD     28     28       POTENTIAL.
FT   CARBOHYD     49     49       POTENTIAL.
SQ   SEQUENCE   238 AA;  25238 MW;  284A50FD CRC32;

  Query Match           63.5%;  Score 47;  DB 1;  Length 238;
  Best Local Similarity  50.0%;  Pred. No. 2.36e+01;
  Matches     5;  Conservative    5;  Mismatches   0;  Indels   0;  Gaps   0;

Db    180 SFLTASPALD 189
          |||::|::|:
Qy      2 SFLSTSSSLE 11


RESULT   15
ID   PH4H_PSEAE     STANDARD;      PRT;   262 AA.
AC   P43334;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PHENYLALANINE-4-HYDROXYLASE (EC 1.14.16.1) (PAH) (PHE-4-
DE   MONOOXYGENASE).
GN   PHHA.
OS   PSEUDOMONAS AERUGINOSA.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC   PSEUDOMONAS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE; 94151331.
RA   ZHAO G., XIA T., SONG J., ROY R.A.;
RT   "Pseudomonas aeruginosa possesses homologues of mammalian
RT   phenylalanine hydroxylase and 4 alpha-carbinolamine dehydratase/DCoH
RT   as part of a three-component gene cluster.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 91:1366-1370(1994).
CC   -!- CATALYTIC ACTIVITY: L-PHENYLALANINE + TETRAHYDROBIOPTERIN + O(2) =
CC       L-TYROSINE + DIHYDROBIOPTERIN + H(2)O.
CC   -!- COFACTOR: THIS ENZYME REQUIRES A FERROUS ION.
CC   -!- PATHWAY: RATE-LIMITING STEP IN PHENYLALANINE CATABOLISM.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC       HYDROXYLASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M88627; G476741; -.
DR   PROSITE; PS00367; BIOPTERIN_HYDROXYL; 1.
DR   PFAM; PF00351; biopterin_H; 1.
DR   HSSP; P04177; 1TOH.
KW   OXIDOREDUCTASE; MONOOXYGENASE; PHENYLALANINE CATABOLISM; IRON.
FT   METAL       121    121       IRON (BY SIMILARITY).
FT   METAL       126    126       IRON (BY SIMILARITY).
SQ   SEQUENCE   262 AA;  30288 MW;  6D464AD4 CRC32;

  Query Match           63.5%;  Score 47;  DB 1;  Length 262;
  Best Local Similarity  41.7%;  Pred. No. 2.36e+01;
  Matches     5;  Conservative    4;  Mismatches   3;  Indels   0;  Gaps   0;

Db    102 ATFIRTPEELDY 113
          |:|: |:  |:|
Qy      1 ASFLSTSSSLEY 12


Search completed: Thu Sep  2 12:47:38 1999
Job time : 8 secs.
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U50987; G1575011; -.
KW   TRICARBOXYLIC ACID CYCLE; ELECTRON TRANSPORT; HEME; TRANSMEMBRANE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   TRANSIT       1     55       MITOCHONDRION (POTENTIAL).
FT   CHAIN        56    158       SUCCINATE DEHYDROGENASE [UBIQUINONE]
FT                                 CYTOCHROME B SMALL SUBUNIT.
FT   TRANSMEM     70     90       POTENTIAL.
FT   TRANSMEM    125    141       POTENTIAL.
SQ   SEQUENCE   158 AA;  17096 MW;  703D5238 CRC32;

  Query Match           64.9%;  Score 48;  DB 1;  Length 158;
  Best Local Similarity  33.3%;  Pred. No. 1.47e+01;
  Matches     4;  Conservative    6;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 AAYLNPCSAMDY 92
          |::|:  |:::|
Qy      1 ASFLSTSSSLEY 12


RESULT   12
ID   DHSD_HUMAN     STANDARD;      PRT;   159 AA.
AC   O14521;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT
DE   PRECURSOR (CYBS) (SUCCINATE-UBIQUINONE REDUCTASE MEMBRANE ANCHOR
DE   SUBUNIT).
GN   SDHD OR SDH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 98194224.
RA   HIRAWAKE H., TANIWAKI M., KIJIMA S., KITA K.;
RT   "Cytochrome b in human complex II (succinate-ubiquinone
RT   oxidoreductase): cDNA cloning of the components in liver mitochondria
RT   and chromosome assignment of the genes for the large (SDHC) and small
RT   (SDHD) subunits to 1q21 and 11q23.";
RL   CYTOGENET. CELL GENET. 79:132-138(1997).
CC   -!- SUBUNIT: COMPOSED OF A 27 KD IRON PROTEIN (IP), A 70 KD
CC       FLAVOPROTEIN (FP) AND A CYTOCHROME B COMPOSED OF TWO INTEGRAL
CC       MEMBRANE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB006202; D1022913; -.
DR   MIM; 602690; -.
KW   TRICARBOXYLIC ACID CYCLE; ELECTRON TRANSPORT; HEME; TRANSMEMBRANE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   TRANSIT       1     56       MITOCHONDRION (POTENTIAL).
FT   CHAIN        57    159       SUCCINATE DEHYDROGENASE [UBIQUINONE]
FT                                 CYTOCHROME B SMALL SUBUNIT.
FT   TRANSMEM     71     91       POTENTIAL.
FT   TRANSMEM    126    142       POTENTIAL.
SQ   SEQUENCE   159 AA;  17043 MW;  F4221825 CRC32;

  Query Match           64.9%;  Score 48;  DB 1;  Length 159;
  Best Local Similarity  33.3%;  Pred. No. 1.47e+01;
  Matches     4;  Conservative    6;  Mismatches   2;  Indels   0;  Gaps   0;

Db     82 AAYLNPCSAMDY 93
          |::|:  |:::|
Qy      1 ASFLSTSSSLEY 12


RESULT   13
ID   YOE1_STRAT     STANDARD;      PRT;   229 AA.
AC   Q53683;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL PROTEIN IN OLED 5'REGION (ORF1) (FRAGMENT).
OS   STREPTOMYCES ANTIBIOTICUS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; STREPTOMYCINEAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11891;
RX   MEDLINE; 94063510.
RA   HERNANDEZ C., OLANO C., MENDEZ C., SALAS J.A.;
RT   "Characterization of a Streptomyces antibioticus gene cluster
RT   encoding a glycosyltransferase involved in oleandomycin
RT   inactivation.";
RL   GENE 134:139-140(1993).
CC   -!- FUNCTION: MAY PARTICIPATE IN OLEANDOMYCIN SECRETION DURING
CC       ANTIBIOTIC PRODUCTION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC       PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE MALFG
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z22577; G296489; -.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; PARTIAL.
DR   PFAM; PF00528; BPD_transp; 1.
KW   HYPOTHETICAL PROTEIN; TRANSMEMBRANE; TRANSPORT.
FT   NON_TER       1      1
FT   TRANSMEM     26     42       POTENTIAL.
FT   TRANSMEM     63     79       POTENTIAL.
FT   TRANSMEM     94    110       POTENTIAL.
FT   TRANSMEM    141    157       POTENTIAL.
FT   TRANSMEM    198    214       POTENTIAL.
SQ   SEQUENCE   229 AA;  24312 MW;  EF5F0F21 CRC32;

  Query Match           63.5%;  Score 47;  DB 1;  Length 229;
  Best Local Similarity  50.0%;  Pred. No. 2.36e+01;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    199 AAFLATLPSLAF 210
          |:||:| :|| :
Qy      1 ASFLSTSSSLEY 12


RESULT   14
ID   VGLG_HSV11     STANDARD;      PRT;   238 AA.
AC   P06484;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   GLYCOPROTEIN G.
GN   GG OR US4.
OS   HERPES SIMPLEX VIRUS (TYPE 1 / STRAIN 17).
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
```

```
  Best Local Similarity  60.0%;  Pred. No. 9.09e+00;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db      8 FLESVSPLEY 17
          || : |:|||
Qy      3 FLSTSSSLEY 12


RESULT    9
ID   CYSP_THEPA     STANDARD;      PRT;   439 AA.
AC   P22497;
DT   01-AUG-1991 (REL. 19, CREATED)
DT   01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CYSTEINE PROTEINASE PRECURSOR (EC 3.4.22.-).
OS   THEILERIA PARVA.
OC   EUKARYOTA; ALVEOLATA; APICOMPLEXA; PIROPLASMIDA; THEILERIIDAE;
OC   THEILERIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91009278.
RA   NENE V., GOBRIGHT E., MUSOKE A.J., LONSDALE-ECCLES J.D.;
RT   "A single exon codes for the enzyme domain of a protozoan cysteine
RT   protease.";
RL   J. BIOL. CHEM. 265:18047-18050(1990).
RN   [2]
RP   SEQUENCE OF 178-439 FROM N.A.
RC   STRAIN=MUGUGA;
RX   MEDLINE; 92228011.
RA   NENE V., IAMS K.P., GOBRIGHT E., MUSOKE A.J.;
RT   "Characterisation of the gene encoding a candidate vaccine antigen of
RT   Theileria parva sporozoites.";
RL   MOL. BIOCHEM. PARASITOL. 51:17-27(1992).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC       PAPAIN FAMILY OF THIOL PROTEASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37791; G161873; -.
DR   EMBL; M67476; G161879; -.
DR   PIR; A36083; KHQBTT.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PFAM; PF00112; Cys-protease; 1.
DR   HSSP; P14080; 1YAL.
KW   HYDROLASE; THIOL PROTEASE; ZYMOGEN; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     59       POTENTIAL.
FT   PROPEP       60    228       ACTIVATION PEPTIDE.
FT   CHAIN       229    439       CYSTEINE PROTEINASE.
FT   ACT_SITE    252    252       BY SIMILARITY.
FT   ACT_SITE    381    381       BY SIMILARITY.
FT   ACT_SITE    403    403       BY SIMILARITY.
FT   DISULFID    249    290       BY SIMILARITY.
FT   DOMAIN      165    181       INVOLVED IN PROCESSING TO YIELD ACTIVE
FT                                ENZYMES (BY SIMILARITY).
FT   CARBOHYD    205    205       POTENTIAL.
SQ   SEQUENCE   439 AA;  50179 MW;  B4BD9301 CRC32;

  Query Match           66.2%;  Score 49;  DB 1;  Length 439;
  Best Local Similarity  54.5%;  Pred. No. 9.09e+00;
  Matches     6;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db    108 GFLSDDPKLEY 118
          :|||  : |||
Qy      2 SFLSTSSSLEY 12


RESULT   10
ID   YA9A_SCHPO     STANDARD;      PRT;   530 AA.
AC   Q09788;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 54.2 KD SERINE-RICH PROTEIN C13G6.10C IN CHROMOSOME I
DE   PRECURSOR.
GN   SPAC13G6.10C.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   ODELL C., BOWMAN S., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (SEP-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z54308; G1008995; -.
KW   HYPOTHETICAL PROTEIN; SIGNAL.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    530       HYPOTHETICAL PROTEIN C13G6.10C.
FT   CARBOHYD     55     55       POTENTIAL.
FT   CARBOHYD    120    120       POTENTIAL.
FT   CARBOHYD    128    128       POTENTIAL.
SQ   SEQUENCE   530 AA;  54210 MW;  48D83906 CRC32;

  Query Match           66.2%;  Score 49;  DB 1;  Length 530;
  Best Local Similarity  54.5%;  Pred. No. 9.09e+00;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     11 AFLSTVGALPY 21
          :|||| ::| |
Qy      2 SFLSTSSSLEY 12


RESULT   11
ID   DHSD_BOVIN     STANDARD;      PRT;   158 AA.
AC   Q95123;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT
DE   PRECURSOR (CYBS) (SUCCINATE-UBIQUINONE REDUCTASE MEMBRANE ANCHOR
DE   SUBUNIT) (QPS3).
GN   SDHD OR SDH4.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RA   SHENOY S.K., YU L., YU C.A.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBUNIT: COMPOSED OF A 27 KD IRON PROTEIN (IP), A 70 KD
CC       FLAVOPROTEIN (FP) AND A CYTOCHROME B COMPOSED OF TWO INTEGRAL
CC       MEMBRANE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
FT   REPEAT       387    422       1.
FT   REPEAT       423    471       2.
FT   REPEAT       472    519       3.
FT   REPEAT       520    566       4.
SQ   SEQUENCE   579 AA;  66042 MW;  60543E32 CRC32;

  Query Match           67.6%;  Score 50;  DB 1;  Length 579;
  Best Local Similarity  50.0%;  Pred. No. 5.57e+00;
  Matches     6;  Conservative    3;  Mismatches   3;  Indels   0;  Gaps   0;

Db    133 AAFLQMSIALDY 144
          |:||  | :|:|
Qy      1 ASFLSTSSSLEY 12


RESULT    7
ID   XMRK_XIPMA     STANDARD;      PRT;  1166 AA.
AC   P13388;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   MELANOMA RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112).
GN   XMRK OR TU.
OS   XIPHOPHORUS MACULATUS (SOUTHERN PLATYFISH).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; ATHERINOMORPHA;
OC   CYPRINODONTIFORMES; CYPRINODONTOIDEI; POECILIIDAE; XIPHOPHORUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90015140.
RA   WITTBRODT J., ADAM D., MALITSCHEK B., MAUELER W., RAULF F.,
RA   TELLING A., ROBERTSON S.M., SCHARTL M.;
RT   "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT   inducing Tu locus in Xiphophorus.";
RL   NATURE 341:415-421(1989).
CC   -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16891; G65291; -.
DR   PIR; S06142; S06142.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00757; Furin-like; 1.
DR   PFAM; PF01030; Recep_L_domain; 2.
DR   HSSP; P11362; 1FGI.
KW   TRANSMEMBRANE; GLYCOPROTEIN; DUPLICATION; RECEPTOR; SIGNAL;
KW   TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; PHOSPHORYLATION;
KW   PROTO-ONCOGENE.
FT   SIGNAL        1     25
FT   CHAIN        26   1166       MELANOMA RECEPTOR TYROSINE KINASE.
FT   DOMAIN       26    642       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    643    665       POTENTIAL.
FT   DOMAIN      666   1166       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      709    976       PROTEIN KINASE.
FT   NP_BIND     715    723       ATP (BY SIMILARITY).
FT   BINDING     742    742       ATP (BY SIMILARITY).
FT   ACT_SITE    834    834       BY SIMILARITY.
FT   CARBOHYD    114    114       POTENTIAL.
FT   CARBOHYD    144    144       POTENTIAL.
FT   CARBOHYD    201    201       POTENTIAL.
FT   CARBOHYD    356    356       POTENTIAL.
FT   CARBOHYD    365    365       POTENTIAL.
FT   CARBOHYD    398    398       POTENTIAL.
FT   CARBOHYD    417    417       POTENTIAL.
FT   CARBOHYD    501    501       POTENTIAL.
FT   CARBOHYD    575    575       POTENTIAL.
FT   CARBOHYD    620    620       POTENTIAL.
SQ   SEQUENCE   1166 AA;  129877 MW;  48F8E7C0 CRC32;

  Query Match           67.6%;  Score 50;  DB 1;  Length 1166;
  Best Local Similarity  50.0%;  Pred. No. 5.57e+00;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1145 FLPAAENLEY 1154
          ||::: :|||
Qy      3 FLSTSSSLEY 12


RESULT    8
ID   RIR2_MYCPN     STANDARD;      PRT;   339 AA.
AC   P75461;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE BETA CHAIN (EC 1.17.4.1)
DE   (RIBONUCLEOTIDE REDUCTASE).
GN   NRDF.
OS   MYCOPLASMA PNEUMONIAE.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC   MYCOPLASMATACEAE; MYCOPLASMA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE; 97105885.
RA   HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA   HERRMANN R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC   -!- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM
CC       THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY
CC       FOR DNA SYNTHESIS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC       THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC       THIOREDOXIN.
CC   -!- COFACTOR: CONTAINS TWO IRON IONS (BY SIMILARITY).
CC   -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC       SMALL CHAIN FAMILY. MORE SIMILAR TO ENTEROBACTERIAL NRDF THAN TO
CC       NRDB. SEEMS TO LACK TWO OF THE IRON-BINDING RESIDUES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000050; G1674208; -.
DR   PROSITE; PS00368; RIBORED_SMALL; FALSE_NEG.
DR   PFAM; PF00268; ribonuc_red; 1.
KW   OXIDOREDUCTASE; DNA REPLICATION; IRON.
FT   METAL         87     87       IRON 1 (BY SIMILARITY).
FT   METAL        121    121       IRON 1 (BY SIMILARITY).
FT   METAL        215    215       IRON 2.
FT   ACT_SITE     125    125       BY SIMILARITY.
SQ   SEQUENCE   339 AA;  39413 MW;  B73F2761 CRC32;

  Query Match           66.2%;  Score 49;  DB 1;  Length 339;
```

```
RESULT    4
ID   UFO2_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16165;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MC2 ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13501; G295854; -.
DR   PIR; S08325; S08325.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48621 MW;  3158C5E0 CRC32;

  Query Match           67.6%;  Score 50;  DB 1;  Length 471;
  Best Local Similarity  88.9%;  Pred. No. 5.57e+00;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     47 SFLSTASSL 55
          |||||:|||
Qy      2 SFLSTSSSL 10


RESULT    5
ID   UFO3_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16167;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-W22 ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88284304.
RA   RALSTON E.J., ENGLISH J.J., DOONER H.K.;
RT   "Sequence of three bronze alleles of maize and correlation with the
RT   genetic fine structure.";
RL   GENETICS 119:185-197(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13502; G22506; -.
DR   EMBL; X07937; G22210; -.
DR   PIR; S01037; S01037.
DR   PIR; S08326; S08326.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48673 MW;  4A3C6193 CRC32;

  Query Match           67.6%;  Score 50;  DB 1;  Length 471;
  Best Local Similarity  88.9%;  Pred. No. 5.57e+00;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     47 SFLSTASSL 55
          |||||:|||
Qy      2 SFLSTSSSL 10


RESULT    6
ID   YR47_CAEEL     STANDARD;      PRT;   579 AA.
AC   Q09563;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 66.0 KD PROTEIN F47D12.7 IN CHROMOSOME III.
GN   F47D12.7.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   TAICH A.;
RL   SUBMITTED (MAR-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: STRONG, TO C.ELEGANS T16H12.6 AND SOME, TO MOUSE MIPP.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U22831; G722359; -.
DR   WORMPEP; F47D12.7; CE01951.
DR   PFAM; PF00651; BTB; 1.
DR   PFAM; PF01344; Kelch; 6.
KW   HYPOTHETICAL PROTEIN; REPEAT.
FT   DOMAIN      387    566       4 X APPROXIMATE TANDEM REPEATS.
```

```
KW   AROMATIC HYDROCARBONS CATABOLISM; OXIDOREDUCTASE; MONOOXYGENASE;
KW   FLAVOPROTEIN; FAD; IRON; PLASMID.
SQ   SEQUENCE   517 AA;  60522 MW;  FB61602A CRC32;

  Query Match           70.3%;  Score 52;  DB 1;  Length 517;
  Best Local Similarity  60.0%;  Pred. No. 2.05e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    101 FLTAVSPLEY 110
          ||:: |:|||
Qy      3 FLSTSSSLEY 12


RESULT    2
ID   WCAM_ECOLI     STANDARD;      PRT;   464 AA.
AC   P71244; P76378;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   COLANIC ACID BIOSYNTHESIS PROTEIN WCAM.
GN   WCAM.
OS   ESCHERICHIA COLI.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ESCHERICHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 96326333.
RA   STEVENSON G., HOBBS M., ANDRIANOPOULOS K., REEVES P.;
RT   "Organization of the Escherichia coli K-12 gene cluster responsible
RT   for production of the extracellular polysaccharide colanic acid.";
RL   J. BACTERIOL. 178:4885-4893(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   SCIENCE 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97251358.
RA   ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T.,
RA   ISONO K., KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M.,
RA   MAKINO K., MIKI T., MIZOBUCHI K., MORI H., MORI T., MOTOMURA K.,
RA   NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,
RA   SAITO N., SAMPEI G., SEKI Y., SIVASUNDARAM S., TAGAMI H.,
RA   TAKEDA J., TAKEMOTO K., WADA C., YAMAMOTO Y., HORIUCHI T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA RES. 3:379-392(1996).
CC   -!- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF THE SLIME POLYSACCHARIDE
CC       COLANIC ACID.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000295; G1788356; -.
DR   EMBL; U38473; G1407622; -.
DR   EMBL; D90842; G1736746; -.
DR   ECOGENE; EG12651; WCAM.
KW   LIPOPOLYSACCHARIDE BIOSYNTHESIS.
FT   CONFLICT      3      3       F -> S (IN REF. 1).
FT   CONFLICT     14     16       ASS -> KL (IN REF. 1).
FT   CONFLICT     83     87       HILGS -> THSWN (IN REF. 1).
FT   CONFLICT    457    464       FRLPELRE -> LLDCRN (IN REF. 1).
SQ   SEQUENCE   464 AA;  51315 MW;  40AF62BE CRC32;

  Query Match           67.6%;  Score 50;  DB 1;  Length 464;
  Best Local Similarity  45.5%;  Pred. No. 5.57e+00;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     10 TFLTASSALAF 20
          :||::||:| :
Qy      2 SFLSTSSSLEY 12


RESULT    3
ID   UFO1_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16166;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MCC ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88284304.
RA   RALSTON E.J., ENGLISH J.J., DOONER H.K.;
RT   "Sequence of three bronze alleles of maize and correlation with the
RT   genetic fine structure.";
RL   GENETICS 119:185-197(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13500; G1030071; -.
DR   EMBL; X07940; G22205; -.
DR   PIR; S01052; S01052.
DR   PIR; S08324; S08324.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48769 MW;  8AE03FD2 CRC32;

  Query Match           67.6%;  Score 50;  DB 1;  Length 471;
  Best Local Similarity  88.9%;  Pred. No. 5.57e+00;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     47 SFLSTASSL 55
          |||||:|||
Qy      2 SFLSTSSSL 10
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

on: Thu Sep 2 12:47:30 1999; MasPar time 2.21 Seconds
153.600 Million cell updates/sec
ular output not generated.

Title: >US-08-599-226-33
Description: (1-12) from US08599226.pep
Perfect Score: 74
Sequence: 1 ASFLSTSSSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 25.085; Variance 27.334; scale 0.918

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	52	70.3	517	1	DMPN_PSESP	PHENOL HYDROXYLASE P3	2.05e+00
2	50	67.6	464	1	WCAM_ECOLI	COLANIC ACID BIOSYNTHE	5.57e+00
3	50	67.6	471	1	UFO1_MAIZE	FLAVONOL 3-O-GLUCOSYLT	5.57e+00
4	50	67.6	471	1	UFO2_MAIZE	FLAVONOL 3-O-GLUCOSYLT	5.57e+00
5	50	67.6	471	1	UFO3_MAIZE	FLAVONOL 3-O-GLUCOSYLT	5.57e+00
6	50	67.6	579	1	YR47_CAEEL	HYPOTHETICAL 66.0 KD P	5.57e+00
7	50	67.6	1166	1	XMRK_XIPMA	MELANOMA RECEPTOR PROT	5.57e+00
8	49	66.2	339	1	RIR2_MYCPN	RIBONUCLEOSIDE-DIPHOSP	9.09e+00
9	49	66.2	439	1	CYSP_THEPA	CYSTEINE PROTEINASE PR	9.09e+00
10	49	66.2	530	1	YA9A_SCHPO	HYPOTHETICAL 54.2 KD S	9.09e+00
11	48	64.9	158	1	DHSD_BOVIN	SUCCINATE DEHYDROGENAS	1.47e+01
12	48	64.9	159	1	DHSD_HUMAN	SUCCINATE DEHYDROGENAS	1.47e+01
13	47	63.5	229	1	YOE1_STRAT	HYPOTHETICAL PROTEIN I	2.36e+01
14	47	63.5	238	1	VGLG_HSV11	GLYCOPROTEIN G.	2.36e+01
15	47	63.5	262	1	PH4H_PSEAE	PHENYLALANINE-4-HYDROX	2.36e+01
16	47	63.5	262	1	COX3_PROWI	CYTOCHROME C OXIDASE P	2.36e+01
17	47	63.5	518	1	YNV6_CAEEL	HYPOTHETICAL 59.0 KD P	2.36e+01
18	47	63.5	563	1	HEMA_IAMAA	HEMAGGLUTININ PRECURSO	2.36e+01
19	47	63.5	568	1	HEMA_IAMAB	HEMAGGLUTININ PRECURSO	2.36e+01
20	47	63.5	792	1	Y85K_SSV1	HYPOTHETICAL 85.7 KD P	2.36e+01
21	47	63.5	828	1	YKR6_YEAST	HYPOTHETICAL 93.4 KD P	2.36e+01
22	46	62.2	207	1	CLPP_ECOLI	ATP-DEPENDENT CLP PROT	3.75e+01
23	46	62.2	218	1	Y232_SYNY3	HYPOTHETICAL 23.8 KD P	3.75e+01
24	46	62.2	237	1	CYPR_DROME	PEPTIDYL-PROLYL CIS-TR	3.75e+01
25	46	62.2	238	1	ARD1_YEAST	N-TERMINAL ACETYLTRANS	3.75e+01
26	46	62.2	270	1	VA32_VARV	PROTEIN A32.	3.75e+01
27	46	62.2	300	1	VA32_VACCV	PROTEIN A32.	3.75e+01
28	46	62.2	360	1	1A1D_ECOLI	PUTATIVE 1-AMINOCYCLOP	3.75e+01
29	46	62.2	395	1	NEUA_STRAG	ACYLNEURAMINATE CYTIDY	3.75e+01
30	46	62.2	411	1	LAG1_YEAST	LONGEVITY-ASSURANCE PR	3.75e+01
31	46	62.2	448	1	Y4QG_RHISN	PROBABLE AMINOTRANSFER	3.75e+01
32	46	62.2	449	1	TBA_SORMA	TUBULIN ALPHA CHAIN.	3.75e+01
33	46	62.2	504	1	PPOM_TOBAC	PROTOPORPHYRINOGEN OXI	3.75e+01
34	46	62.2	505	1	MATK_SAXIN	PROBABLE INTRON MATURA	3.75e+01
35	46	62.2	543	1	YJF8_YEAST	HYPOTHETICAL 60.8 KD P	3.75e+01
36	46	62.2	773	1	LI19_DROME	LIN-19 HOMOLOG PROTEIN	3.75e+01
37	46	62.2	779	1	SRP_DROME	BOX A-BINDING FACTOR (	3.75e+01
38	46	62.2	882	1	HSS1_RAT	HEPARAN SULFATE N-DEAC	3.75e+01
39	46	62.2	917	1	GLRK_LYMST	GLUTAMATE RECEPTOR PRE	3.75e+01
40	46	62.2	926	1	KINH_NEUCR	KINESIN HEAVY CHAIN.	3.75e+01
41	46	62.2	1038	1	SOG_DROME	DORSAL-VENTRAL PATTERN	3.75e+01
42	46	62.2	1164	1	YHV8_YEAST	HYPOTHETICAL 131.1 KD	3.75e+01
43	45	60.8	214	1	RB17_MOUSE	RAS-RELATED PROTEIN RA	5.91e+01
44	45	60.8	257	1	YGIE_ECOLI	HYPOTHETICAL 26.5 KD P	5.91e+01
45	45	60.8	1003	1	PUR2_CHICK	PHOSPHORIBOSYLAMINE--G	5.91e+01

ALIGNMENTS

```
RESULT    1
ID   DMPN_PSESP     STANDARD;      PRT;   517 AA.
AC   P19732;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PHENOL HYDROXYLASE P3 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE
DE   P3 COMPONENT).
GN   DMPN OR PHEA4.
OS   PSEUDOMONAS SP. (STRAIN CF600).
OG   PLASMID PVI150.
OC   BACTERIA; PROTEOBACTERIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91072230.
RA   NORDLUND I., POWLOWSKI J., SHINGLER V.;
RT   "Complete nucleotide sequence and polypeptide analysis of
RT   multicomponent phenol hydroxylase from Pseudomonas sp. strain
RT   CF600.";
RL   J. BACTERIOL. 172:6826-6833(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BH;
RA   TAKEO M., MAEDA Y., OKADA H., MIYAMA K., MORI K., IKE M.,
RA   FUJITA M.;
RL   SUBMITTED (MAR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC       DERIVATIVES. P3 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR
CC       IN VITRO PHENOL HYDROXYLASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+)
CC       + H(2)O.
CC   -!- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY.
CC   -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC   -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED
CC       BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M60276; G151453; -.
DR   EMBL; D28864; G468469; -.
DR   PIR; D37831; D37831.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.


```
MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

Run on:          Thu Sep  2 12:51:38 1999;  MasPar time 4.47 Seconds
                                            146.610 Million cell updates/sec
Tabular output not generated.

Title:           >US-08-599-226-34
Description:     (1-12) from US08599226.pep
Perfect Score:   75
Sequence:        1 ASYLSTASSLEY 12

Scoring table:   PAM 150
                 Gap 15

Searched:        179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:        sptrembl9
                   1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
                   5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
                   9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
                   13:sp_vertebrate 14:sp_virus

Statistics:      Mean 24.013;  Variance 31.833;  scale 0.754

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	58	77.3	649	3	O60167	PROTEIN COMPLEX ASSEMB	7.07e-01
2	57	76.0	503	2	O26074	PROTEIN-EXPORT MEMBRAN	1.13e+00
3	53	70.7	897	5	Q17336	LET 858.	6.94e+00
4	52	69.3	469	2	O32204	YVSH PROTEIN.	1.08e+01
5	52	69.3	582	3	O74931	ALTERNATIVE NADH-DEHYD	1.08e+01
6	52	69.3	1238	5	O61198	F15E6.6 PROTEIN.	1.08e+01
7	51	68.0	293	2	O51095	CONSERVED HYPOTHETICAL	1.67e+01
8	51	68.0	349	5	O17959	M01B2.5 PROTEIN.	1.67e+01
9	51	68.0	475	13	O93514	AXIAL PROTOCADHERIN (F	1.67e+01
10	51	68.0	932	5	O01623	SIMILAR TO LIGAND-GATE	1.67e+01
11	50	66.7	83	9	O48385	ORF83.	2.56e+01
12	50	66.7	127	14	O11696	NUCLEOPROTEIN (FRAGMEN	2.56e+01
13	50	66.7	222	13	P70023	OLFACTORY RECEPTOR (FR	2.56e+01
14	50	66.7	265	5	Q25144	HROX1.	2.56e+01
15	50	66.7	372	5	P91143	SIMILAR TO ACETYLTRANS	2.56e+01
16	50	66.7	1186	14	O55767	PUTATIVE TYROSINE PROT	2.56e+01
17	49	65.3	122	10	P93302	ORF122A.	3.91e+01
18	49	65.3	383	2	Q56675	MANNOSE-SENSITIVE HEMA	3.91e+01
19	49	65.3	910	3	O59796	PUTATIVE RECEPTOR ASSO	3.91e+01
20	49	65.3	939	4	O15071	KIAA0365 (FRAGMENT).	3.91e+01
21	49	65.3	949	4	O60369	KIAA0365 (FRAGMENT).	3.91e+01
22	49	65.3	1406	4	O15082	KIAA0377.	3.91e+01
23	48	64.0	290	14	Q98138	GENE 27 PROTEIN.	5.93e+01
24	48	64.0	311	13	Q99162	MELANOMA RECEPTOR PROT	5.93e+01
25	48	64.0	343	14	O40927	ORF 27.	5.93e+01
26	48	64.0	347	2	P95504	PCB1, PCB2 AND PCB3 GE	5.93e+01
27	48	64.0	350	2	P95503	CHLOROPHYLL A/B ANTENN	5.93e+01
28	48	64.0	357	1	O73958	PRIMASE SMALL SUBUNIT	5.93e+01
29	48	64.0	377	2	O84397	HYPOTHETICAL 41.4 KD P	5.93e+01
30	48	64.0	392	2	O50873	CONSERVED HYPOTHETICAL	5.93e+01
31	48	64.0	409	2	O25853	NADH-UBIQUINONE OXIDOR	5.93e+01
32	48	64.0	444	5	P91141	SIMILAR TO ACETYLTRANS	5.93e+01
33	48	64.0	471	10	O23552	HYPOTHETICAL 52.8 KD P	5.93e+01
34	48	64.0	511	2	Q43981	PHENOLHYDROXYLASE COMP	5.93e+01
35	48	64.0	626	2	Q46732	COLICIN PROTEIN.	5.93e+01
36	48	64.0	626	2	Q46738	COLICIN PROTEIN.	5.93e+01
37	48	64.0	626	2	Q46734	COLICIN PROTEIN.	5.93e+01
38	48	64.0	626	2	Q46740	COLICIN PROTEIN.	5.93e+01
39	48	64.0	771	3	O42812	PROLYL DIPEPTIDYL PEPT	5.93e+01
40	48	64.0	989	11	Q63052	APOLIPOPROTEIN B (FRAG	5.93e+01
41	48	64.0	1026	4	Q08174	PROTOCADHERIN 42 PRECU	5.93e+01
42	48	64.0	1035	13	O57537	NF-PROTOCADHERIN.	5.93e+01
43	48	64.0	1069	11	O88185	BH-PROTOCADHERIN-A.	5.93e+01
44	48	64.0	1069	4	O60245	PCDH7 (BH-PCDH)A.	5.93e+01
45	48	64.0	1200	4	O60247	PCDH7 (BH-PCDH)C.	5.93e+01

ALIGNMENTS

```
RESULT    1
ID   O60167      PRELIMINARY;      PRT;   649 AA.
AC   O60167;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   PROTEIN COMPLEX ASSEMBLY PROTEIN.
GN   SPBC19F8.03C.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   BECK A., REINHARDT R., WOOD V., RAJANDREAM M.A., BARRELL B.G.;
RL   SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AL023594; E1293401; -.
SQ   SEQUENCE   649 AA;  72985 MW;  9C207DB2 CRC32;

  Query Match             77.3%;  Score 58;  DB 3;  Length 649;
  Best Local Similarity   72.7%;  Pred. No. 7.07e-01;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Db     234 NYLSTARSLEF 244
           :||||| |||:
Qy       2 SYLSTASSLEY 12


RESULT    2
ID   O26074      PRELIMINARY;      PRT;   503 AA.
AC   O26074;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PROTEIN-EXPORT MEMBRANE PROTEIN (SECD).
GN   HP1550.
OS   HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC   BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC   HELICOBACTER.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695;
RX   MEDLINE; 97394467.
```

```
RA   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori[published erratum appears in Nature 1997 Sep
RT   25;389(6649):412].";
RL   NATURE 388:539-547(1997).
DR   EMBL; AE000652; G2314730; -.
DR   TIGR; HP1550; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   503 AA;  54247 MW;  9A76592C CRC32;

  Query Match           76.0%;  Score 57;  DB 2;  Length 503;
  Best Local Similarity  80.0%;  Pred. No. 1.13e+00;
  Matches     8;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db     38 YLSLASALEY 47
          ||| ||:|||
Qy      3 YLSTASSLEY 12


RESULT    3
ID   Q17336      PRELIMINARY;      PRT;   897 AA.
AC   Q17336;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   LET 858.
GN   LET-858.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2 (BRISTOL);
RA   KELLY W.G., COLES L.H., FIRE A.Z.;
RL   GENETICS 0:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MATTHEWS L.;
RL   SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U19615; G987227; -.
DR   EMBL; Z81525; E1351661; -.
     SEQUENCE   897 AA;  104268 MW;  E1E3EA36 CRC32;

  Query Match           70.7%;  Score 53;  DB 5;  Length 897;
  Best Local Similarity  60.0%;  Pred. No. 6.94e+00;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    491 YLTMQSSLDY 500
          ||:  |||:|
Qy      3 YLSTASSLEY 12


RESULT    4
ID   O32204      PRELIMINARY;      PRT;   469 AA.
AC   O32204;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   YVSH PROTEIN.
GN   YVSH.
OS   BACILLUS SUBTILIS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA   AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA   BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA   BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA   CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA   DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA   ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA   FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA   GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA   GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA   HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA   JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA   KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA   KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA   LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA   MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA   NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA   PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA   PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA   RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA   SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA   SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
RA   SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA   TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA   TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA   VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA   WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA   YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   NATURE 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   WIPAT A., BRIGNELL C.S., GUY J.B., ROSE M., EMMERSON P.T.,
RA   HARWOOD C.R.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z99121; E1186022; -.
DR   EMBL; AJ223978; E1249790; -.
DR   PFAM; PF00324; aa_permeases; 1.
SQ   SEQUENCE   469 AA;  50258 MW;  49186162 CRC32;

  Query Match           69.3%;  Score 52;  DB 2;  Length 469;
  Best Local Similarity  45.5%;  Pred. No. 1.08e+01;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    376 TFLTTAATLAY 386
          ::|:||::| |
Qy      2 SYLSTASSLEY 12


RESULT    5
ID   O74931      PRELIMINARY;      PRT;   582 AA.
AC   O74931;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ALTERNATIVE NADH-DEHYDROGENASE PRECURSOR (EC 1.6.5.3)
DE   (NADH DEHYDROGENASE (UBIQUINONE)) (UBIQUINONE REDUCTASE)
DE   (TYPE I DEHYDROGENASE) (COMPLEX I DEHYDROGENASE).
GN   NDH2.
OS   YARROWIA LIPOLYTICA (CANDIDA LIPOLYTICA).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   DIPODASCACEAE; YARROWIA.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=E150;
RA   KERSCHER S.J., BRANDT U.;
RT   "Identification of the YLNDH2 Gene Encoding the Alternative
RT   NADH:Ubiquinone Oxidoreductase from Yarrowia lipolytica.";
RL   SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   -!- COFACTOR: FAD; IRON-SULFUR.
DR   EMBL; AJ006852; E1330342; -.
KW   SIGNAL; OXIDOREDUCTASE.
FT   SIGNAL        1     79       POTENTIAL.
FT   CHAIN        80    582       ALTERNATIVE NADH-DEHYDROGENASE.
SQ   SEQUENCE   582 AA;  65814 MW;  0460C796 CRC32;

  Query Match           69.3%;  Score 52;  DB 3;  Length 582;
  Best Local Similarity  54.5%;  Pred. No. 1.08e+01;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 TYLSAIASLGY 91
          :|||: :|| |
Qy      2 SYLSTASSLEY 12


RESULT    6
ID   O61198      PRELIMINARY;      PRT;  1238 AA.
AC   O61198;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F15E6.6 PROTEIN.
GN   F15E6.6.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   MILLER N., STELLYES L., BRADSHAW H., KEPPLER D.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF038614; G2702437; -.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW   IRON-SULFUR.
SQ   SEQUENCE   1238 AA;  135726 MW;  9408BB7C CRC32;

  Query Match           69.3%;  Score 52;  DB 5;  Length 1238;
  Best Local Similarity  70.0%;  Pred. No. 1.08e+01;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    702 YLETQSSLNY 711
          || | |||:|
Qy      3 YLSTASSLEY 12


RESULT    7
ID   O51095      PRELIMINARY;      PRT;  293 AA.
AC   O51095;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   BB0068.
OS   BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC   BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE; 98065943.
RA   FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA   LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA   DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA   PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA   VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA   UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA   GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA   SMITH H.O., VENTER J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   NATURE 390:580-586(1997).
DR   EMBL; AE001120; G2687956; -.
DR   TIGR; BB0068; -.
SQ   SEQUENCE   293 AA;  33278 MW;  3FBBB9E2 CRC32;

  Query Match           68.0%;  Score 51;  DB 2;  Length 293;
  Best Local Similarity  70.0%;  Pred. No. 1.67e+01;
  Matches     7;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    198 AYLSTPNSLE 207
          :||||::|||
Qy      2 SYLSTASSLE 11


RESULT    8
ID   O17959      PRELIMINARY;      PRT;  349 AA.
AC   O17959;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   M01B2.5 PROTEIN.
GN   M01B2.5.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LLOYD C.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z83116; E1348127; -.
SQ   SEQUENCE   349 AA;  40017 MW;  90870FE2 CRC32;
```

```
  Query Match           68.0%;  Score 51;  DB 5;  Length 349;
  Best Local Similarity  50.0%;  Pred. No. 1.67e+01;
  Matches    5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    198 YLENSSGLDY 207
          ||  :|:|:|
Qy      3 YLSTASSLEY 12


RESULT    9
ID   O93514      PRELIMINARY;      PRT;   475 AA.
AC   O93514;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   AXIAL PROTOCADHERIN (FRAGMENT).
GN   AXPC.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YAMAMOTO A., DEROBERTIS E.M.;
RT   "Xenopus axial protocadherin.";
RL   SUBMITTED (MAR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF053469; G3598688; -.
FT   NON_TER       1      1
FT   NON_TER     475    475
SQ   SEQUENCE   475 AA;  52268 MW;  2A681544 CRC32;

  Query Match           68.0%;  Score 51;  DB 13;  Length 475;
  Best Local Similarity  50.0%;  Pred. No. 1.67e+01;
  Matches    5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    341 FLQTTTSLDY 350
          :| |::||:|
Qy      3 YLSTASSLEY 12


RESULT   10
ID   O01623      PRELIMINARY;      PRT;   932 AA.
AC   O01623;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO LIGAND-GATED IONIC CHANNEL PROTEINS.
GN   ZC196.7.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   MURRAY J.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U97007; G1938466; -.
DR   PFAM; PF00060; lig_chan; 2.
SQ   SEQUENCE   932 AA;  106836 MW;  469F21FA CRC32;

  Query Match           68.0%;  Score 51;  DB 5;  Length 932;
  Best Local Similarity  54.5%;  Pred. No. 1.67e+01;
  Matches    6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    753 AYLMESTSLEY 763
          :||  ::||||
Qy      2 SYLSTASSLEY 12


RESULT   11
ID   O48385      PRELIMINARY;      PRT;    83 AA.
AC   O48385;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ORF83.
OS   STREPTOCOCCUS THERMOPHILUS BACTERIOPHAGE TP-J34.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TP-J34;
RX   MEDLINE; 98122991.
RA   NEVE H., ZENZ K.I., DESIERE F., KOCH A., HELLER K.J., BRUSSOW H.;
RT   "Comparison of the lysogeny modules from the temperate Streptococcus
RT   thermophilus bacteriophages TP-J34 and Sfi21: implications for the
RT   modular theory of phage evolution.";
RL   VIROLOGY 241:61-72(1998).
DR   EMBL; AF020798; G2897100; -.
SQ   SEQUENCE   83 AA;  9876 MW;  250B3F9C CRC32;

  Query Match           66.7%;  Score 50;  DB 9;  Length 83;
  Best Local Similarity  60.0%;  Pred. No. 2.56e+01;
  Matches    6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db     71 YLETSAFLEY 80
          || |:: |||
Qy      3 YLSTASSLEY 12


RESULT   12
ID   O11696      PRELIMINARY;      PRT;   127 AA.
AC   O11696;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NUCLEOPROTEIN (FRAGMENT).
OS   MEASLES VIRUS (SUBACUTE SCLEROSE PANENCEPHALITIS VIRUS).
OC   VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC   PARAMYXOVIRIDAE; PARAMYXOVIRINAE; MORBILLIVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=92-E;
RX   MEDLINE; 97278133.
RA   YAMAGUCHI S.;
RT   "Identification of three lineages of wild measles virus by nucleotide
RT   sequence analysis of N, P, M, F, and L genes in Japan.";
RL   J. MED. VIROL. 52:113-120(1997).
DR   EMBL; D87487; D1020995; -.
KW   NUCLEOPROTEIN.
FT   NON_TER       1      1
SQ   SEQUENCE   127 AA;  13950 MW;  42D75A2C CRC32;

  Query Match           66.7%;  Score 50;  DB 14;  Length 127;
  Best Local Similarity  45.5%;  Pred. No. 2.56e+01;
  Matches    5;  Conservative    6;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db     66 AAYLPTSTPLD 76
          |:||:|:::|:
Qy      1 ASYLSTASSLE 11


RESULT   13
ID   P70023      PRELIMINARY;      PRT;   222 AA.
AC   P70023;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   OLFACTORY RECEPTOR (FRAGMENT).
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FREITAG J.;
RL   SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96112032.
RA   FREITAG J., KRIEGER J., STROTMAN J., BREER H.;
RT   "Two classes of olfactory receptors in Xenopus laevis.";
RL   NEURON 15:1383-1392(1995).
DR   EMBL; Y08352; E273933; -.
DR   PFAM; PF00001; 7tm_1; 1.
FT   NON_TER       1      1
FT   NON_TER     222    222
SQ   SEQUENCE   222 AA;  24658 MW;  F18F14F8 CRC32;

  Query Match           66.7%;  Score 50;  DB 13;  Length 222;
  Best Local Similarity  70.0%;  Pred. No. 2.56e+01;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db     96 YASTISSLDY 105
          | || |||:|
Qy      3 YLSTASSLEY 12


RESULT   14
ID   Q25144      PRELIMINARY;      PRT;   265 AA.
AC   Q25144;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HROX1.
GN   HROX1.
OS   HALIOTIS RUFESCENS (CALIFORNIA RED ABALONE).
OC   EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PROSOBRANCHIA;
OC   ARCHAEOGASTROPODA; HALIOTIDAE; HALIOTIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97388321.
RA   DEGNAN B.M., DEGNAN S.M., FENTENANY G., MORSE D.E.;
RT   "A Mox homeobox gene in the gastropod mollusc Haliotis rufescens is
RT   differentially expressed during larval morphogenesis and
RT   metamorphosis.";
RL   FEBS LETT. 411:119-122(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; X75217; G407415; -.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PFAM; PF00046; homeobox; 1.
KW   HOMEOBOX; DNA-BINDING; NUCLEAR PROTEIN.
SQ   SEQUENCE   265 AA;  29579 MW;  B68A753D CRC32;

  Query Match           66.7%;  Score 50;  DB 5;  Length 265;
  Best Local Similarity  58.3%;  Pred. No. 2.56e+01;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 GSYLSMSSSKDY 92
          :|||| :|| :|
Qy      1 ASYLSTASSLEY 12


RESULT   15
ID   P91143      PRELIMINARY;      PRT;   372 AA.
AC   P91143;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO ACETYLTRANSFERASES.
GN   C37H5.2.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DAVIDSON S., GILLAM B.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U88315; G1825777; -.
DR   PFAM; PF00561; abhydrolase; 1.
KW   TRANSFERASE.
SQ   SEQUENCE   372 AA;  42139 MW;  5214F159 CRC32;

  Query Match           66.7%;  Score 50;  DB 5;  Length 372;
  Best Local Similarity  50.0%;  Pred. No. 2.56e+01;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    168 GGYLSTSYALKY 179
          ::||||: :| |
Qy      1 ASYLSTASSLEY 12


Search completed: Thu Sep  2 12:52:11 1999
Job time : 33 secs.
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:51:12 1999; MasPar time 2.19 Seconds
155.131 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-34
Description: (1-12) from US08599226.pep
Perfect Score: 75
Sequence: 1 ASYLSTASSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 25.133; Variance 27.887; scale 0.901

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	51	68.0	1038	1	SOG_DROME	DORSAL-VENTRAL PATTERN	4.04e+00
2	50	66.7	158	1	DHSD_BOVIN	SUCCINATE DEHYDROGENAS	6.58e+00
3	50	66.7	159	1	DHSD_HUMAN	SUCCINATE DEHYDROGENAS	6.58e+00
4	50	66.7	428	1	B4AR_MELGA	BETA-4C ADRENERGIC REC	6.58e+00
5	50	66.7	517	1	DMPN_PSESP	PHENOL HYDROXYLASE P3	6.58e+00
6	50	66.7	564	1	5HT1_DROME	5-HYDROXYTRYPTAMINE RE	6.58e+00
7	49	65.3	352	1	YDH3_SCHPO	HYPOTHETICAL 39.7 KD P	1.06e+01
8	49	65.3	441	1	SECY_MYCTU	PREPROTEIN TRANSLOCASE	1.06e+01
9	49	65.3	475	1	YM61_YEAST	HYPOTHETICAL 55.3 KD P	1.06e+01
10	49	65.3	917	1	GLRK_LYMST	GLUTAMATE RECEPTOR PRE	1.06e+01
11	48	64.0	272	1	YIGL_HAEIN	HYPOTHETICAL PROTEIN H	1.70e+01
12	48	64.0	471	1	UFO1_MAIZE	FLAVONOL 3-O-GLUCOSYLT	1.70e+01
13	48	64.0	471	1	UFO2_MAIZE	FLAVONOL 3-O-GLUCOSYLT	1.70e+01
14	48	64.0	471	1	UFO3_MAIZE	FLAVONOL 3-O-GLUCOSYLT	1.70e+01
15	48	64.0	626	1	CEIB_ECOLI	COLICIN IB PROTEIN.	1.70e+01
16	48	64.0	1166	1	XMRK_XIPMA	MELANOMA RECEPTOR PROT	1.70e+01
17	48	64.0	1255	1	ERB2_HUMAN	ERBB-2 RECEPTOR PROTEI	1.70e+01
18	47	62.7	37	1	YRYL_CAEEL	HYPOTHETICAL 4.1 KD PR	2.71e+01
19	47	62.7	127	1	MBP_RAT	MYELIN BASIC PROTEIN S	2.71e+01
20	47	62.7	167	1	MBP_CAVPO	MYELIN BASIC PROTEIN (	2.71e+01
21	47	62.7	171	1	MBP_PANTR	MYELIN BASIC PROTEIN (	2.71e+01
22	47	62.7	194	1	MBP_MOUSE	MYELIN BASIC PROTEIN (	2.71e+01
23	47	62.7	196	1	MBP_HUMAN	MYELIN BASIC PROTEIN (	2.71e+01
24	47	62.7	209	1	YCCR_ECOLI	HYPOTHETICAL 24.1 KD P	2.71e+01
25	47	62.7	330	1	ODBA_BACSU	2-OXOISOVALERATE DEHYD	2.71e+01
26	47	62.7	339	1	RIR2_MYCPN	RIBONUCLEOSIDE-DIPHOSP	2.71e+01
27	47	62.7	399	1	HXAA_MOUSE	HOMEOBOX PROTEIN HOX-A	2.71e+01
28	47	62.7	485	1	G6PD_ZYMMO	GLUCOSE-6-PHOSPHATE 1-	2.71e+01
29	47	62.7	530	1	YA9A_SCHPO	HYPOTHETICAL 54.2 KD S	2.71e+01
30	47	62.7	577	1	PYRH_YEAST	CTP SYNTHASE 2 (EC 6.3	2.71e+01
31	47	62.7	653	1	YJU4_YEAST	HYPOTHETICAL 75.9 KD P	2.71e+01
32	47	62.7	684	1	YV18_MYCTU	HYPOTHETICAL 57.3 KD P	2.71e+01
33	47	62.7	697	1	TRP_SCHPO	TRYPTOPHAN SYNTHASE (E	2.71e+01
34	47	62.7	848	1	AMPN_LACLA	AMINOPEPTIDASE N (EC 3	2.71e+01
35	46	61.3	301	1	P34_RICRI	PROTEIN P34.	4.27e+01
36	46	61.3	341	1	CG2A_DAUCA	G2/MITOTIC-SPECIFIC CY	4.27e+01
37	46	61.3	413	1	DCHS_LYCES	HISTIDINE DECARBOXYLAS	4.27e+01
38	46	61.3	416	1	YEIJ_ECOLI	HYPOTHETICAL 43.4 KD P	4.27e+01
39	46	61.3	483	1	K1CL_MOUSE	KERATIN, TYPE I CYTOSK	4.27e+01
40	46	61.3	505	1	RADA_SYNY3	DNA REPAIR PROTEIN RAD	4.27e+01
41	46	61.3	554	1	NBL4_MOUSE	NBL4 PROTEIN.	4.27e+01
42	46	61.3	753	1	YBHJ_ECOLI	HYPOTHETICAL 81.5 KD P	4.27e+01
43	46	61.3	824	1	NPA2_HUMAN	NEURONAL PAS DOMAIN PR	4.27e+01
44	46	61.3	1023	1	RT12_ACTPL	RTX-I TOXIN DETERMINAN	4.27e+01
45	46	61.3	1469	1	DP27_CAEEL	CHROMOSOME CONDENSATIO	4.27e+01

ALIGNMENTS

```
RESULT    1
ID   SOG_DROME      STANDARD;      PRT;  1038 AA.
AC   Q24025;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   DORSAL-VENTRAL PATTERNING PROTEIN SOG (SHORT GASTRULATION PROTEIN).
GN   SOG.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95047398.
RA   FRANCOIS V., SOLLOWAY M., O'NEILL J.W., EMERY J., BIER E.;
RT   "Dorsal-ventral patterning of the Drosophila embryo depends on a
RT   putative negative growth factor encoded by the short gastrulation
RT   gene.";
RL   GENES DEV. 8:2602-2616(1994).
CC   -!- FUNCTION: PUTATIVE NEGATIVE GROWTH FACTOR; ANTAGONIST OF DPP, A
CC       PROTEIN INVOLVED IN PATTERNING THE DORSAL REGION AND IN THE
CC       DEVELOPMENT OF THE NEUROECTODERM.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: ABUTS THE DORSAL DPP-EXPRESSING CELLS IN A
CC       LATERAL STRIPE 14-16 CELLS WIDE. LATER IN EMBRYOGENESIS IT IS
CC       EXPRESSED IN NEUROECTODERM AND IN THE ENDODERM SPACED ALONG THE
CC       ANTERIOR-POSTERIOR AXIS OF THE DEVELOPING GUT.
CC   -!- DEVELOPMENTAL STAGE: EMBRYOGENESIS.
CC   -!- SIMILARITY: TO XENOPUS DORSALIZING FACTOR CHORDIN.
CC   -!- SIMILARITY: CONTAINS 4 VWFC DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18774; G1203794; -.
DR   FLYBASE; FBgn0003463; sog.
DR   PROSITE; PS01208; VWFC; 2.
DR   PFAM; PF00093; vwc; 4.
KW   TRANSMEMBRANE; DEVELOPMENTAL PROTEIN; REPEAT; GROWTH FACTOR;
KW   GROWTH REGULATION; SIGNAL-ANCHOR.
FT   DOMAIN        1     53       CYTOPLASMIC (POTENTIAL).
```

```
FT   TRANSMEM      54     74       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                 (POTENTIAL).
FT   DOMAIN        75   1038       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       100    175       VWFC 1.
FT   REPEAT       421    522       SR1.
FT   REPEAT       592    668       SR2.
FT   REPEAT       677    754       SR3.
FT   DOMAIN       742    804       VWFC 2.
FT   DOMAIN       830    899       VWFC 3.
FT   DOMAIN       939   1020       VWFC 4.
FT   CARBOHYD     179    179       POTENTIAL.
FT   CARBOHYD     287    287       POTENTIAL.
FT   CARBOHYD     520    520       POTENTIAL.
FT   CARBOHYD     666    666       POTENTIAL.
FT   CARBOHYD     752    752       POTENTIAL.
FT   CARBOHYD     821    821       POTENTIAL.
SQ   SEQUENCE   1038 AA;  115514 MW;  DC4DAFF5 CRC32;

  Query Match           68.0%;  Score 51;  DB 1;  Length 1038;
  Best Local Similarity  60.0%;  Pred. No. 4.04e+00;
  Matches    6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    497 YLNTDGSLAY 506
          ||:| :|| |
Qy      3 YLSTASSLEY 12


RESULT    2
ID   DHSD_BOVIN     STANDARD;      PRT;   158 AA.
AC   Q95123;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT
DE   PRECURSOR (CYBS) (SUCCINATE-UBIQUINONE REDUCTASE MEMBRANE ANCHOR
DE   SUBUNIT) (QPS3).
GN   SDHD OR SDH4.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RA   SHENOY S.K., YU L., YU C.A.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBUNIT: COMPOSED OF A 27 KD IRON PROTEIN (IP), A 70 KD
CC       FLAVOPROTEIN (FP) AND A CYTOCHROME B COMPOSED OF TWO INTEGRAL
CC       MEMBRANE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U50987; G1575011; -.
KW   TRICARBOXYLIC ACID CYCLE; ELECTRON TRANSPORT; HEME; TRANSMEMBRANE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   TRANSIT       1     55       MITOCHONDRION (POTENTIAL).
FT   CHAIN        56    158       SUCCINATE DEHYDROGENASE [UBIQUINONE]
FT                                 CYTOCHROME B SMALL SUBUNIT.
FT   TRANSMEM      70     90       POTENTIAL.
FT   TRANSMEM     125    141       POTENTIAL.
SQ   SEQUENCE   158 AA;  17096 MW;  703D5238 CRC32;

  Query Match           66.7%;  Score 50;  DB 1;  Length 158;
  Best Local Similarity  41.7%;  Pred. No. 6.58e+00;
  Matches    5;  Conservative    5;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 AAYLNPCSAMDY 92
          |:||:  |:::|
Qy      1 ASYLSTASSLEY 12


RESULT    3
ID   DHSD_HUMAN     STANDARD;      PRT;   159 AA.
AC   O14521;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT
DE   PRECURSOR (CYBS) (SUCCINATE-UBIQUINONE REDUCTASE MEMBRANE ANCHOR
DE   SUBUNIT).
GN   SDHD OR SDH4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 98194224.
RA   HIRAWAKE H., TANIWAKI M., KIJIMA S., KITA K.;
RT   "Cytochrome b in human complex II (succinate-ubiquinone
RT   oxidoreductase): cDNA cloning of the components in liver mitochondria
RT   and chromosome assignment of the genes for the large (SDHC) and small
RT   (SDHD) subunits to 1q21 and 11q23.";
RL   CYTOGENET. CELL GENET. 79:132-138(1997).
CC   -!- SUBUNIT: COMPOSED OF A 27 KD IRON PROTEIN (IP), A 70 KD
CC       FLAVOPROTEIN (FP) AND A CYTOCHROME B COMPOSED OF TWO INTEGRAL
CC       MEMBRANE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB006202; D1022913; -.
DR   MIM; 602690; -.
KW   TRICARBOXYLIC ACID CYCLE; ELECTRON TRANSPORT; HEME; TRANSMEMBRANE;
KW   MITOCHONDRION; TRANSIT PEPTIDE.
FT   TRANSIT       1     56       MITOCHONDRION (POTENTIAL).
FT   CHAIN        57    159       SUCCINATE DEHYDROGENASE [UBIQUINONE]
FT                                 CYTOCHROME B SMALL SUBUNIT.
FT   TRANSMEM      71     91       POTENTIAL.
FT   TRANSMEM     126    142       POTENTIAL.
SQ   SEQUENCE   159 AA;  17043 MW;  F4221825 CRC32;

  Query Match           66.7%;  Score 50;  DB 1;  Length 159;
  Best Local Similarity  41.7%;  Pred. No. 6.58e+00;
  Matches    5;  Conservative    5;  Mismatches   2;  Indels   0;  Gaps   0;

Db     82 AAYLNPCSAMDY 93
          |:||:  |:::|
Qy      1 ASYLSTASSLEY 12


RESULT    4
ID   B4AR_MELGA     STANDARD;      PRT;   428 AA.
AC   P43141;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   BETA-4C ADRENERGIC RECEPTOR.
GN   ADRB4C.
OS   MELEAGRIS GALLOPAVO (COMMON TURKEY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; MELEAGRIDIDAE; MELEAGRIS.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95014249.
RA   CHEN X.-H., HARDEN T.K., NICHOLAS R.A.;
RT   "Molecular cloning and characterization of a novel beta-adrenergic
RT   receptor.";
RL   J. BIOL. CHEM. 269:24810-24819(1994).
CC   -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC       INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC       PROTEINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: BROAD TISSUE DISTRIBUTION.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U13977; G556604; -.
DR   EMBL; U13978; G555882; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
DR   HSSP; P07700; 1DEP.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT   DOMAIN        1     25       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     26     49       1 (POTENTIAL).
FT   DOMAIN       50     58       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     59     77       2 (POTENTIAL).
FT   DOMAIN       78     97       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     98    119       3 (POTENTIAL).
FT   DOMAIN      120    141       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    142    164       4 (POTENTIAL).
FT   DOMAIN      165    189       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    190    211       5 (POTENTIAL).
FT   DOMAIN      212    261       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    262    283       6 (POTENTIAL).
FT   DOMAIN      284    294       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    295    315       7 (POTENTIAL).
FT   DOMAIN      316    428       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      8      8       POTENTIAL.
FT   CARBOHYD     13     13       POTENTIAL.
FT   DISULFID     96    175       BY SIMILARITY.
FT   LIPID       329    329       PALMITATE (BY SIMILARITY).
SQ   SEQUENCE   428 AA;  47398 MW;  8B794F0C CRC32;

  Query Match           66.7%;  Score 50;  DB 1;  Length 428;
  Best Local Similarity  40.0%;  Pred. No. 6.58e+00;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    122 YLAITAPLQY 131
          ||: :::|:|
Qy      3 YLSTASSLEY 12


RESULT    5
ID   DMPN_PSESP     STANDARD;      PRT;   517 AA.
AC   P19732;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PHENOL HYDROXYLASE P3 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE
DE   P3 COMPONENT).
GN   DMPN OR PHEA4.
OS   PSEUDOMONAS SP. (STRAIN CF600).
OG   PLASMID PVI150.
OC   BACTERIA; PROTEOBACTERIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91072230.
RA   NORDLUND I., POWLOWSKI J., SHINGLER V.;
RT   "Complete nucleotide sequence and polypeptide analysis of
RT   multicomponent phenol hydroxylase from Pseudomonas sp. strain
RT   CF600.";
RL   J. BACTERIOL. 172:6826-6833(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BH;
RA   TAKEO M., MAEDA Y., OKADA H., MIYAMA K., MORI K., IKE M.,
RA   FUJITA M.;
RL   SUBMITTED (MAR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC       DERIVATIVES. P3 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR
CC       IN VITRO PHENOL HYDROXYLASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+)
CC       + H(2)O.
CC   -!- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY.
CC   -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC   -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED
CC       BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M60276; G151453; -.
DR   EMBL; D28864; G468469; -.
DR   PIR; D37831; D37831.
KW   AROMATIC HYDROCARBONS CATABOLISM; OXIDOREDUCTASE; MONOOXYGENASE;
KW   FLAVOPROTEIN; FAD; IRON; PLASMID.
SQ   SEQUENCE   517 AA;  60522 MW;  FB61602A CRC32;

  Query Match           66.7%;  Score 50;  DB 1;  Length 517;
  Best Local Similarity  50.0%;  Pred. No. 6.58e+00;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    101 FLTAVSPLEY 110
          :|:: |:|||
Qy      3 YLSTASSLEY 12


RESULT    6
ID   5HT1_DROME     STANDARD;      PRT;   564 AA.
AC   P20905;
DT   01-FEB-1991 (REL. 17, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   5-HYDROXYTRYPTAMINE RECEPTOR 1 (5-HT RECEPTOR) (SEROTONIN RECEPTOR).
GN   5HT-R1 OR 5-HT7.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=HEAD;
RX   MEDLINE; 91062395.
RA   WITZ P., AMLAIKY N., PLASSAT J.-L., MAROTEAUX L., BORRELLI E., HEN R.;
RT   "Cloning and characterization of a Drosophila serotonin receptor that
RT   activates adenylate cyclase.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 87:8940-8944(1990).
CC   -!- FUNCTION: THIS IS ONE OF THE SEVERAL DIFFERENT RECEPTORS FOR
CC       5-HYDROXYTRYPTAMINE (SEROTONIN), A BIOGENIC HORMONE THAT FUNCTION
CC       AS A NEUROTRANSMITTER, A HORMONE, AND A MITOGEN. THE ACTIVITY OF
CC       THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLATE
CC       CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: HEAD.
```

```
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       BUT WITH ONE EXTRA POTENTIAL TRANSMEMBRANE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M55533; G156725; -.
DR   PIR; A38271; A38271.
DR   GCRDB; GCR_0023; -.
DR   FLYBASE; FBgn0004573; 5-HT7.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT.
FT   TRANSMEM     29     51       0 (POTENTIAL).
FT   TRANSMEM    165    188       1 (POTENTIAL).
FT   DOMAIN      189    198       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    199    222       2 (POTENTIAL).
FT   DOMAIN      223    236       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    237    258       3 (POTENTIAL).
FT   DOMAIN      259    278       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    279    302       4 (POTENTIAL).
FT   DOMAIN      303    330       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    331    353       5 (POTENTIAL).
FT   DOMAIN      354    454       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    455    476       6 (POTENTIAL).
FT   DOMAIN      477    487       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    488    510       7 (POTENTIAL).
FT   DOMAIN      511    564       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       89    106       9 X 2 AA TANDEM REPEATS OF G-S.
FT   DISULFID    235    314       BY SIMILARITY.
SQ   SEQUENCE   564 AA;  60861 MW;  312369E8 CRC32;

  Query Match           66.7%;  Score 50;  DB 1;  Length 564;
  Best Local Similarity  50.0%;  Pred. No. 6.58e+00;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    261 YLAITKPLEY 270
          ||: : :|||
Qy      3 YLSTASSLEY 12


RESULT    7
ID   YDH3_SCHPO     STANDARD;      PRT;   352 AA.
AC   Q92348;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 39.7 KD PROTEIN C6G9.03C IN CHROMOSOME I.
GN   SPAC6G9.03C.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., CONNOR R.E.;
RL   SUBMITTED (OCT-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: SOME, TO YEAST YNL206C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z81317; E276610; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   352 AA;  39679 MW;  D92A9357 CRC32;

  Query Match           65.3%;  Score 49;  DB 1;  Length 352;
  Best Local Similarity  80.0%;  Pred. No. 1.06e+01;
  Matches     8;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    303 SYLNEASSLE 312
          |||: |||||
Qy      2 SYLSTASSLE 11


RESULT    8
ID   SECY_MYCTU     STANDARD;      PRT;   441 AA.
AC   P94926;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PREPROTEIN TRANSLOCASE SECY SUBUNIT.
GN   SECY OR RV0732 OR MTV041.06.
OS   MYCOBACTERIUM TUBERCULOSIS, AND MYCOBACTERIUM BOVIS.
OC   BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC   ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.TUBERCULOSIS; STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   NATURE 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.BOVIS; STRAIN=BCG;
RA   KIM J.K., CHOE Y.K.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECA AND SECE
CC       TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,
CC       BY FORMING PART OF A CHANNEL (BY SIMILARITY).
CC   -!- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
CC       COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE SECY/SEC61-ALPHA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021958; E1253270; -.
DR   EMBL; U77912; G1813627; -.
DR   PROSITE; PS00755; SECY_1; 1.
DR   PROSITE; PS00756; SECY_2; 1.
DR   PFAM; PF00344; secY; 1.
KW   PROTEIN TRANSPORT; TRANSMEMBRANE; TRANSLOCATION.
FT   TRANSMEM     18     38       POTENTIAL.
FT   TRANSMEM     57     77       POTENTIAL.
FT   TRANSMEM     78     98       POTENTIAL.
FT   TRANSMEM    124    144       POTENTIAL.
FT   TRANSMEM    157    177       POTENTIAL.
FT   TRANSMEM    180    200       POTENTIAL.
```

```
FT   TRANSMEM     215    235       POTENTIAL.
FT   TRANSMEM     272    292       POTENTIAL.
FT   TRANSMEM     318    338       POTENTIAL.
FT   TRANSMEM     382    402       POTENTIAL.
FT   TRANSMEM     404    424       POTENTIAL.
SQ   SEQUENCE   441 AA;  47611 MW;  26E1FAC6 CRC32;

  Query Match           65.3%;  Score 49;  DB 1;  Length 441;
  Best Local Similarity  50.0%;  Pred. No. 1.06e+01;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    309 GTYLSDPSNLVY 320
          ::||| :|:| |
Qy      1 ASYLSTASSLEY 12


RESULT    9
ID   YM61_YEAST     STANDARD;      PRT;   475 AA.
AC   Q03652;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 55.3 KD PROTEIN IN RAR1-SCJ1 INTERGENIC REGION.
GN   YMR211W OR YM8261.05.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   DEDMAN K., BROWN D., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.,
RA   WALSH S.V.;
RL   SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49809; G854463; -.
DR   PFAM; PF00091; tubulin; 1.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   475 AA;  55312 MW;  18FA4F03 CRC32;

  Query Match           65.3%;  Score 49;  DB 1;  Length 475;
  Best Local Similarity  60.0%;  Pred. No. 1.06e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    321 YLTTAITLGY 330
          ||:|| :| |
Qy      3 YLSTASSLEY 12


RESULT   10
ID   GLRK_LYMST     STANDARD;      PRT;   917 AA.
AC   P26591;
DT   01-AUG-1992 (REL. 23, CREATED)
DT   01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT   01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE   GLUTAMATE RECEPTOR PRECURSOR.
OS   LYMNAEA STAGNALIS (GREAT POND SNAIL).
OC   EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PULMONATA; BASOMMATOPHORA;
OC   LYMNAEIDAE; LYMNAEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92070466.
RA   HUTTON M.L., HARVEY R.J., BARNARD E.A., DARLISON M.G.;
RT   "Cloning of a cDNA that encodes an invertebrate glutamate receptor
RT   subunit.";
RL   FEBS LETT. 292:111-114(1991).
CC   -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT
CC       MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC
CC       ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60086; G9629; -.
DR   PIR; S15681; ACGAE.
DR   PIR; S18443; S18443.
DR   PFAM; PF00060; lig_chan; 1.
KW   RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
KW   TRANSMEMBRANE.
FT   SIGNAL         1     19       POTENTIAL.
FT   CHAIN         20    917       GLUTAMATE RECEPTOR.
FT   DOMAIN        21    558       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     559    578       POTENTIAL.
FT   TRANSMEM     599    617       POTENTIAL.
FT   TRANSMEM     628    646       POTENTIAL.
FT   TRANSMEM     819    839       POTENTIAL.
FT   CARBOHYD      62     62       POTENTIAL.
FT   CARBOHYD      95     95       POTENTIAL.
FT   CARBOHYD     121    121       POTENTIAL.
FT   CARBOHYD     125    125       POTENTIAL.
FT   CARBOHYD     229    229       POTENTIAL.
FT   CARBOHYD     251    251       POTENTIAL.
FT   CARBOHYD     261    261       POTENTIAL.
FT   CARBOHYD     272    272       POTENTIAL.
FT   CARBOHYD     418    418       POTENTIAL.
FT   CARBOHYD     419    419       POTENTIAL.
FT   CARBOHYD     424    424       POTENTIAL.
FT   CARBOHYD     491    491       POTENTIAL.
FT   CARBOHYD     775    775       POTENTIAL.
FT   CARBOHYD     881    881       POTENTIAL.
SQ   SEQUENCE   917 AA;  103139 MW;  879CBEDC CRC32;

  Query Match           65.3%;  Score 49;  DB 1;  Length 917;
  Best Local Similarity  36.4%;  Pred. No. 1.06e+01;
  Matches     4;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db    729 AYLTESSTIDY 739
          :||: :|:::|
Qy      2 SYLSTASSLEY 12


RESULT   11
ID   YIGL_HAEIN     STANDARD;      PRT;   272 AA.
AC   P44771;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL PROTEIN HI0597.
GN   HI0597.
OS   HAEMOPHILUS INFLUENZAE.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC   HAEMOPHILUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
```

```
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   SCIENCE 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO THE COF/YBHA/YIDA/YIGL (E.COLI) / YCSE/YXEH
CC       (B.SUBTILIS) FAMILY. STRONG, TO E.COLI YIGL.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32741; G1573586; -.
DR   TIGR; HI0597; -.
DR   PROSITE; PS01228; COF_1; 1.
DR   PROSITE; PS01229; COF_2; 1.
DR   PFAM; PF00592; DUF3; 1.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   272 AA;  30523 MW;  7F53B65C CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 272;
  Best Local Similarity  50.0%;  Pred. No. 1.70e+01;
  Matches     6;  Conservative    2;  Mismatches   4;  Indels   0;  Gaps   0;

Db    261 ARYLRTQFGLDY 272
          | || |  :|:|
Qy      1 ASYLSTASSLEY 12


RESULT   12
ID   UFO1_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16166;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MCC ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88284304.
RA   RALSTON E.J., ENGLISH J.J., DOONER H.K.;
RT   "Sequence of three bronze alleles of maize and correlation with the
RT   genetic fine structure.";
RL   GENETICS 119:185-197(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13500; G1030071; -.
DR   EMBL; X07940; G22205; -.
DR   PIR; S01052; S01052.
DR   PIR; S08324; S08324.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48769 MW;  8AE03FD2 CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 471;
  Best Local Similarity  88.9%;  Pred. No. 1.70e+01;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     47 SFLSTASSL 55
          |:|||||||
Qy      2 SYLSTASSL 10


RESULT   13
ID   UFO2_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16165;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-MC2 ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13501; G295854; -.
DR   PIR; S08325; S08325.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48621 MW;  3158C5E0 CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 471;
  Best Local Similarity  88.9%;  Pred. No. 1.70e+01;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     47 SFLSTASSL 55
          |:|||||||
Qy      2 SYLSTASSL 10
```

```
RESULT   14
ID   UFO3_MAIZE     STANDARD;      PRT;   471 AA.
AC   P16167;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   FLAVONOL 3-O-GLUCOSYLTRANSFERASE (EC 2.4.1.91) (UDP-GLUCOSE FLAVONOID
DE   3-O-GLUCOSYLTRANSFERASE) (BRONZE-1) (BZ-W22 ALLELE).
GN   BZ1 OR UGT71A1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC   POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FURTEK D., SCHIEFELBEIN J.W., JOHNSTON F., NELSON O.E. JR.;
RT   "Sequence comparisons of 3 wild-type bronze-1 alleles from Zea mays.";
RL   PLANT MOL. BIOL. 11:473-481(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88284304.
RA   RALSTON E.J., ENGLISH J.J., DOONER H.K.;
RT   "Sequence of three bronze alleles of maize and correlation with the
RT   genetic fine structure.";
RL   GENETICS 119:185-197(1988).
CC   -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
CC       GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
CC       PIGMENTS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL = UDP + FLAVONOL
CC       3-O-D-GLUCOSIDE.
CC   -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
CC       ANTHOCYANIN BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X13502; G22506; -.
DR   EMBL; X07937; G22210; -.
DR   PIR; S01037; S01037.
DR   PIR; S08326; S08326.
DR   MAIZEDB; 13885; -.
DR   PROSITE; PS00375; UDPGT; 1.
DR   PFAM; PF00201; UDPGT; 2.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE.
SQ   SEQUENCE   471 AA;  48673 MW;  4A3C6193 CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 471;
  Best Local Similarity  88.9%;  Pred. No. 1.70e+01;
  Matches     8;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     47 SFLSTASSL 55
          |:|||||||
Qy      2 SYLSTASSL 10


RESULT   15
ID   CEIB_ECOLI     STANDARD;      PRT;   626 AA.
AC   P04479;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   COLICIN IB PROTEIN.
GN   CIB.
OS   ESCHERICHIA COLI.
OG   PLASMID INCI1 COLIB-P9.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ESCHERICHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87008385.
RA   MANKOVICH J.A., HSU C.-H., KONISKY J.;
RT   "DNA and amino acid sequence analysis of structural and immunity
RT   genes of colicins Ia and Ib.";
RL   J. BACTERIOL. 168:228-236(1986).
RN   [2]
RP   SEQUENCE OF 1-40 FROM N.A.
RX   MEDLINE; 84264487.
RA   MANKOVICH J.A., LAI P.H., GOKUL N., KONISKY J.;
RT   "Organization of the colicin Ib gene. Promoter structure and immunity
RT   domain.";
RL   J. BIOL. CHEM. 259:8764-8768(1984).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85014128.
RA   VARLEY J.M., BOULNOIS G.J.;
RT   "Analysis of a cloned colicin Ib gene: complete nucleotide sequence
RT   and implications for regulation of expression.";
RL   NUCLEIC ACIDS RES. 12:6727-6739(1984).
RN   [4]
RP   ERRATUM.
RA   VARLEY J.M., BOULNOIS G.J.;
RL   NUCLEIC ACIDS RES. 12:8748-8748(1984).
CC   -!- FUNCTION: THIS COLICIN IS A CHANNEL-FORMING COLICIN. THIS CLASS OF
CC       TRANSMEMBRANE TOXINS DEPOLARIZE THE CYTOPLASMIC MEMBRANE, LEADING
CC       TO DISSIPATION OF CELLULAR ENERGY.
CC   -!- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
CC       AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
CC   -!- SIMILARITY: BELONGS TO THE CHANNEL FORMING COLICIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; K02071; G144640; -.
DR   EMBL; X01009; G41142; -.
DR   EMBL; M13820; G144648; -.
DR   PIR; A03503; IKECB.
DR   PIR; A22503; A22503.
DR   PIR; D25035; D25035.
DR   PROSITE; PS00276; CHANNEL_COLICIN; 1.
DR   PFAM; PF01024; Colicin; 1.
DR   HSSP; P06716; 1CII.
KW   PLASMID; BACTERIOCIN; COLICIN; TOXIN; TRANSMEMBRANE.
FT   TRANSMEM    588    612       POTENTIAL.
SQ   SEQUENCE   626 AA;  69923 MW;  983D11B0 CRC32;

  Query Match           64.0%;  Score 48;  DB 1;  Length 626;
  Best Local Similarity  54.5%;  Pred. No. 1.70e+01;
  Matches     6;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    333 NYLLTHSGLDY 343
          :|| | |:|:|
Qy      2 SYLSTASSLEY 12


Search completed: Thu Sep  2 12:51:21 1999
Job time : 9 secs.
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:50:40 1999; MasPar time 3.12 Seconds
154.181 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-34
Description: (1-12) from US08599226.pep
Perfect Score: 75
Sequence: 1 ASYLSTASSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.614; Variance 30.527; scale 0.806

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	57	76.0	503	2	F64713	protein-export membra	6.18e-01
2	57	76.0	526	2	D71805	protein-export membra	6.18e-01
3	52	69.3	469	2	D70048	ABC transporter (amin	6.57e+00
4	51	68.0	293	2	D70108	conserved hypothetica	1.03e+01
5	51	68.0	477	2	F71918	hypothetical protein	1.03e+01
6	50	66.7	105	2	S69755	hypothetical protein	1.62e+01
7	50	66.7	265	2	S38380	Hrox1 protein - Calif	1.62e+01
8	50	66.7	428	2	A55044	beta-4C-adrenergic re	1.62e+01
9	50	66.7	517	2	D37831	phenol 2-monooxygenas	1.62e+01
10	50	66.7	564	2	A38271	serotonin receptor 7	1.62e+01
11	50	66.7	1186	2	T03180	tyrosine protein kina	1.62e+01
12	49	65.3	441	2	G70822	probable secY protein	2.52e+01
13	49	65.3	475	2	S55093	hypothetical protein	2.52e+01
14	49	65.3	917	1	ACGAE	glutamate receptor pr	2.52e+01
15	49	65.3	949	3	T03030	hypothetical protein	2.52e+01
16	48	64.0	272	2	D64155	hypothetical protein	3.89e+01
17	48	64.0	296	2	S45336	finger protein, Spl/e	3.89e+01
18	48	64.0	311	2	S13808	protein-tyrosine kina	3.89e+01
19	48	64.0	346	2	S13807	protein-tyrosine kina	3.89e+01
20	48	64.0	346	2	S13809	protein-tyrosine kina	3.89e+01
21	48	64.0	377	2	F71520	hypothetical protein	3.89e+01
22	48	64.0	392	2	B70242	conserved hypothetica	3.89e+01
23	48	64.0	409	2	G64677	NADH dehydrogenase (u	3.89e+01
24	48	64.0	409	2	E71838	nadh oxidoreductase I	3.89e+01
25	48	64.0	471	2	S08325	flavonol 3-O-glucosyl	3.89e+01
26	48	64.0	471	2	S01052	flavonol 3-O-glucosyl	3.89e+01
27	48	64.0	471	2	S01037	flavonol 3-O-glucosyl	3.89e+01
28	48	64.0	471	2	C71439	hypothetical protein	3.89e+01
29	48	64.0	511	2	S47290	phenol 2-monooxygenas	3.89e+01
30	48	64.0	626	1	IKECB	colicin Ib - Escheric	3.89e+01
31	48	64.0	989	2	I56333	apolipoprotein B - ra	3.89e+01
32	48	64.0	1069	2	T00043	BH-protocadherin-a -	3.89e+01
33	48	64.0	1069	2	T00040	BH-protocadherin PCDH	3.89e+01
34	48	64.0	1072	2	T00041	BH-protocadherin PCDH	3.89e+01
35	48	64.0	1166	2	S06142	kinase-related transf	3.89e+01
36	48	64.0	1200	2	T00042	BH-protocadherin PCDH	3.89e+01
37	48	64.0	1255	1	A24571	protein-tyrosine kina	3.89e+01
38	47	62.7	112	4	S59333	hypothetical protein	5.95e+01
39	47	62.7	167	2	A37246	myelin basic protein	5.95e+01
40	47	62.7	180	2	S65026	finger protein XFG 68	5.95e+01
41	47	62.7	209	2	F64836	probable membrane pro	5.95e+01
42	47	62.7	339	2	S73840	ribonucleotide reduct	5.95e+01
43	47	62.7	814	2	I39627	nicotine dehydrogenas	5.95e+01
44	47	62.7	840	2	S77615	hypothetical protein	5.95e+01
45	47	62.7	1746	2	S19694	tenascin precursor -	5.95e+01

ALIGNMENTS

RESULT 1
ENTRY F64713 #type complete
TITLE protein-export membrane protein - Helicobacter pylori (strain 26695)
ORGANISM #formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 12-Feb-1999
ACCESSIONS F64713
REFERENCE A64520
#authors Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen Helicobacter pylori.
#cross-references MUID:97394467
#accession F64713
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-503 ##label TOM
##cross-references GB:AE000652; GB:AE000511; NID:g2314720; PID:g2314730; TIGR:HP1550
GENETICS
#start_codon GTG
CLASSIFICATION #superfamily protein export membrane protein secD
SUMMARY #length 503 #molecular-weight 54247 #checksum 3320

Query Match 76.0%; Score 57; DB 2; Length 503;
Best Local Similarity 80.0%; Pred. No. 6.18e-01;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Db      38 YLSLASALEY 47
           ||| ||:|||
Qy       3 YLSTASSLEY 12
```

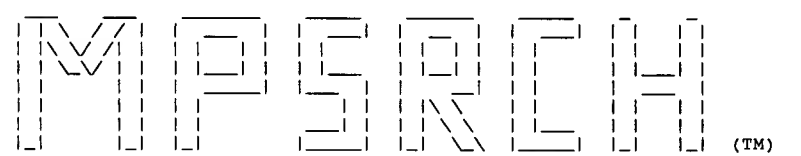
RESULT 2
ENTRY D71805 #type complete

```
TITLE           protein-export membrane protein - Helicobacter pylori (strain
                  J99)
ORGANISM        #formal_name Helicobacter pylori
   #variety     strain J99
DATE            12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                  05-Mar-1999
ACCESSIONS      D71805
REFERENCE       A71800
   #authors     Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                  Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                  B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                  Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                  Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                  G.F.; Trust, T.J.
   #journal     Nature (1999) 397:176-180
   #title       Genomic sequence comparison of two unrelated isolates of the
                  human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession   D71805
      ##status       preliminary
      ##molecule_type DNA
      ##residues     1-526 ##label ARN
      ##cross-references GB:AE001567; GB:AE001439; NID:g4156065; PID:g4156069
      ##experimental_source strain J99
GENETICS
   #gene        secD
CLASSIFICATION  #superfamily protein export membrane protein secD
SUMMARY         #length 526  #molecular-weight 56796  #checksum 5813

  Query Match           76.0%;  Score 57;  DB 2;  Length 526;
  Best Local Similarity  80.0%;  Pred. No. 6.18e-01;
  Matches     8;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db      60 YLSLASALEY 69
           ||| ||:|||
Qy       3 YLSTASSLEY 12


RESULT    3
ENTRY           D70048     #type complete
TITLE           ABC transporter (amino acid permease) homolog yvsH - Bacillus
                  subtilis
ORGANISM        #formal_name Bacillus subtilis
DATE            05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                  24-Sep-1998
ACCESSIONS      D70048
REFERENCE       A69580
   #authors     Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                  Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                  Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                  A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                  Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                  Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                  Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                  Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                  Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                  M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                  S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                  Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                  C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                  Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                  Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                  Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                  Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                  Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                  Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                  M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                  M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                  V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                  A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                  Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                  Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                  Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                  Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                  B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                  Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                  Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                  Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                  Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                  Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                  K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                  Yoshikawa, H.; Danchin, A.
   #journal     Nature (1997) 390:249-256
   #title       The complete genome sequence of the Gram-positive bacterium
                  Bacillus subtilis.
   #cross-references MUID:98044033
   #accession   D70048
      ##status       preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues     1-469 ##label KUN
      ##cross-references GB:Z99121; GB:AL009126; NID:g2635827; PID:e1186022;
                       PID:g2635847
      ##experimental_source strain 168
GENETICS
   #gene        yvsH
CLASSIFICATION  #superfamily L-lysine transport protein
SUMMARY         #length 469  #molecular-weight 50258  #checksum 4200

  Query Match           69.3%;  Score 52;  DB 2;  Length 469;
  Best Local Similarity  45.5%;  Pred. No. 6.57e+00;
  Matches     5;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db     376 TFLTTAATLAY 386
           ::|:||::| |
Qy       2 SYLSTASSLEY 12


RESULT    4
ENTRY           D70108     #type complete
TITLE           conserved hypothetical protein BB0068 - Lyme disease
                  spirochete
ORGANISM        #formal_name Borrelia burgdorferi #common_name Lyme disease
                  spirochete
DATE            13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                  05-Jun-1998
ACCESSIONS      D70108
REFERENCE       A70100
   #authors     Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
                  Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
                  Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
                  J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
                  Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
                  Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
                  J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
                  Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
                  K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
   #journal     Nature (1997) 390:580-586
   #title       Genomic sequence of a Lyme disease spirochaete, Borrelia
                  burgdorferi.
   #cross-references MUID:98065943
   #accession   D70108
      ##status       preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues     1-293 ##label KLE
      ##cross-references GB:AE001120; GB:AE000783; NID:g2687951; PID:g2687956;
                       TIGR:BB0068
      ##experimental_source strain B31
SUMMARY         #length 293  #molecular-weight 33278  #checksum 5223

  Query Match           68.0%;  Score 51;  DB 2;  Length 293;
  Best Local Similarity  70.0%;  Pred. No. 1.03e+01;
  Matches     7;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db     198 AYLSTPNSLE 207
```

```
          :||||::|||
Qy      2 SYLSTASSLE 11


RESULT    5
ENTRY             F71918     #type complete
TITLE             hypothetical protein jhp0552 - Helicobacter pylori (strain
                    J99)
ORGANISM          #formal_name Helicobacter pylori
   #variety       strain J99
DATE              12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                    12-Feb-1999
ACCESSIONS        F71918
REFERENCE         A71800
   #authors       Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                    Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                    B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                    Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                    Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                    G.F.; Trust, T.J.
   #journal       Nature (1999) 397:176-180
   #title         Genomic sequence comparison of two unrelated isolates of the
                    human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession     F71918
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-477 ##label ARN
      ##cross-references GB:AE001487; GB:AE001439; NID:g4155086; PID:g4155090
      ##experimental_source strain J99
GENETICS
   #gene          jhp0552
SUMMARY           #length 477  #molecular-weight 54625  #checksum 3264

  Query Match             68.0%;  Score 51;  DB 2;  Length 477;
  Best Local Similarity   58.3%;  Pred. No. 1.03e+01;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    172 ATYRSTVANLEY 183
          |:| || ::|||
Qy      1 ASYLSTASSLEY 12


RESULT    6
ENTRY             S69755     #type complete
TITLE             hypothetical protein YDR203w - yeast (Saccharomyces
                    cerevisiae)
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              23-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
                    12-Dec-1997
ACCESSIONS        S69755
REFERENCE         S52697
   #authors       Oliver, K.; Harris, D.
   #submission    submitted to the EMBL Data Library, March 1995
   #accession     S69755
      ##molecule_type DNA
      ##residues      1-105 ##label OLI
      ##cross-references EMBL:Z48784; MIPS:YDR203w
GENETICS
   #map_position 4R
SUMMARY           #length 105  #molecular-weight 11606  #checksum 5265

  Query Match             66.7%;  Score 50;  DB 2;  Length 105;
  Best Local Similarity   54.5%;  Pred. No. 1.62e+01;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db     74 SYLSSAPNLSF 84
          ||||:|::| :
Qy      2 SYLSTASSLEY 12


RESULT    7
ENTRY             S38380     #type complete
TITLE             Hrox1 protein - California red abalone
ORGANISM          #formal_name Haliotis rufescens #common_name California red
                    abalone
DATE              13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
                    17-Oct-1997
ACCESSIONS        S38380
REFERENCE         S38380
   #authors       Degnan, B.M.; Degnan, S.M.; Morse, D.E.
   #submission    submitted to the EMBL Data Library, September 1993
   #description   Expression of Hrox, a gastropod mollusc Hox homeobox gene, is
                    progressively restricted during morphogenesis from
                    trochophore to veliger larval forms.
   #accession     S38380
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-265 ##label DEG
      ##cross-references EMBL:X75217; NID:g407414; PID:g407415
CLASSIFICATION    #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS          DNA binding; homeobox; nucleus; transcription regulation
FEATURE
   164-220             #domain homeobox homology #label HOX
SUMMARY           #length 265  #molecular-weight 29579  #checksum 8996

  Query Match             66.7%;  Score 50;  DB 2;  Length 265;
  Best Local Similarity   58.3%;  Pred. No. 1.62e+01;
  Matches     7;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 GSYLSMSSSKDY 92
          :|||| :|| :|
Qy      1 ASYLSTASSLEY 12


RESULT    8
ENTRY             A55044     #type complete
TITLE             beta-4C-adrenergic receptor - turkey
ORGANISM          #formal_name Meleagris gallopavo #common_name common turkey
DATE              18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change
                    08-Sep-1997
ACCESSIONS        A55044
REFERENCE         A55044
   #authors       Chen, X.; Harden, T.K.; Nicholas, R.A.
   #journal       J. Biol. Chem. (1994) 269:24810-24819
   #title         Molecular cloning and characterization of a novel
                    beta-adrenergic receptor.
   #accession     A55044
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-428 ##label CHE
      ##cross-references GB:U13978; NID:g555881; PID:g555882
GENETICS
   #introns       416/2
CLASSIFICATION    #superfamily vertebrate rhodopsin
KEYWORDS          neurotransmitter receptor; transmembrane protein
SUMMARY           #length 428  #molecular-weight 47398  #checksum 8085

  Query Match             66.7%;  Score 50;  DB 2;  Length 428;
  Best Local Similarity   40.0%;  Pred. No. 1.62e+01;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    122 YLAITAPLQY 131
          ||: :::|:|
Qy      3 YLSTASSLEY 12


RESULT    9
ENTRY             D37831     #type complete
TITLE             phenol 2-monooxygenase (EC 1.14.13.7) chain P3 - Pseudomonas
                    sp. (strain CF600)
ORGANISM          #formal_name Pseudomonas sp.
DATE              14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change
                    31-Oct-1997
ACCESSIONS        D37831
REFERENCE         A37831
```

```
   #authors       Nordlund, I.; Powlowski, J.; Shingler, V.
   #journal       J. Bacteriol. (1990) 172:6826-6833
   #title         Complete nucleotide sequence and polypeptide analysis of
                    multicomponent phenol hydroxylase from Pseudomonas sp.
                    strain CF600.
   #cross-references MUID:91072230
   #accession     D37831
      ##status          preliminary
      ##molecule_type DNA
      ##residues       1-517 ##label NOR
      ##cross-references GB:M60276; GB:M37764; NID:g151449; PID:g151453
KEYWORDS          oxidoreductase
SUMMARY           #length 517  #molecular-weight 60522  #checksum 6342

  Query Match           66.7%;  Score 50;  DB 2;  Length 517;
  Best Local Similarity  50.0%;  Pred. No. 1.62e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    101 FLTAVSPLEY 110
          :|:: |:|||
Qy      3 YLSTASSLEY 12


RESULT   10
ENTRY             A38271     #type complete
TITLE             serotonin receptor 7 - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES   5-hydroxytryptamine receptor 7 (5-HTR7)
ORGANISM          #formal_name Drosophila melanogaster
DATE              22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change
                    04-Sep-1998
ACCESSIONS        A38271
REFERENCE         A38271
   #authors       Witz, P.; Amlaiky, N.; Plassat, J.L.; Maroteaux, L.;
                    Borrelli, E.; Hen, R.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8940-8944
   #title         Cloning and characterization of a Drosophila serotonin
                    receptor that activates adenylate cyclase.
   #cross-references MUID:91062395
   #accession     A38271
      ##molecule_type mRNA
      ##residues       1-564 ##label WIT
      ##cross-references GB:M55533
      ##note          the authors translated the codon CTT for residue 213 as
                       Ala, GTG for residue 215 as Leu, ATG for residue 280
                       as Leu and ATG for residue 551 as Asn
GENETICS
   #gene          FlyBase:5-HT7
      ##cross-references FlyBase:FBgn0004573
CLASSIFICATION    #superfamily octopamine receptor type I
KEYWORDS          G protein-coupled receptor; glycoprotein; membrane protein;
                    phosphoprotein
SUMMARY           #length 564  #molecular-weight 60934  #checksum 3324

  Query Match           66.7%;  Score 50;  DB 2;  Length 564;
  Best Local Similarity  50.0%;  Pred. No. 1.62e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    261 YLAITKPLEY 270
          ||: : :|||
Qy      3 YLSTASSLEY 12


RESULT   11
ENTRY             T03180     #type complete
TITLE             tyrosine protein kinase homolog - Chilo iridescent virus
ORGANISM          #formal_name Chilo iridescent virus
DATE              24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
                    24-Mar-1999
ACCESSIONS        T03180
REFERENCE         Z14834
   #authors       Bahr, U.; Tidona, C.A.; Darai, G.
   #journal       Virus Genes (1997) 15:235-245
   #title         The DNA sequence of Chilo iridescent virus between the genome
                    coordinates 0.101 and 0.391; similarities in coding
                    strategy between insect and vertebrate iridoviruses.
   #accession     T03180
      ##status          preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-1186 ##label BAH
      ##cross-references EMBL:AF003534; NID:g2738385; PID:g2738451
SUMMARY           #length 1186  #molecular-weight 138020  #checksum 6252

  Query Match           66.7%;  Score 50;  DB 2;  Length 1186;
  Best Local Similarity  54.5%;  Pred. No. 1.62e+01;
  Matches     6;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db    503 TYLSRESILDY 513
          :|||  | |:|
Qy      2 SYLSTASSLEY 12


RESULT   12
ENTRY             G70822     #type complete
TITLE             probable secY protein - Mycobacterium tuberculosis (strain
                    H37RV)
ORGANISM          #formal_name Mycobacterium tuberculosis
DATE              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                    21-Nov-1998
ACCESSIONS        G70822
REFERENCE         A70500
   #authors       Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                    Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal       Nature (1998) 393:537-544
   #title         Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
   #cross-references MUID:98295987
   #accession     G70822
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-441 ##label COL
      ##cross-references GB:AL021958; GB:AL123456; NID:g3261536; PID:e1253270;
                         PID:g2911006
      ##experimental_source strain H37Rv
GENETICS
   #gene          secY
CLASSIFICATION    #superfamily preprotein translocase secY
SUMMARY           #length 441  #molecular-weight 47611  #checksum 3743

  Query Match           65.3%;  Score 49;  DB 2;  Length 441;
  Best Local Similarity  50.0%;  Pred. No. 2.52e+01;
  Matches     6;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db    309 GTYLSDPSNLVY 320
          ::||| :|:| |
Qy      1 ASYLSTASSLEY 12


RESULT   13
ENTRY             S55093     #type complete
TITLE             hypothetical protein YMR211w - yeast (Saccharomyces
                    cerevisiae)
ALTERNATE_NAMES   hypothetical protein YM8261.05
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
                    21-Nov-1997
ACCESSIONS        S55093
REFERENCE         S55089
```

```
   #authors      Dedman, K.; Brown, D.; Bowman, S.
   #submission   submitted to the EMBL Data Library, June 1995
   #accession    S55093
      ##molecule_type DNA
      ##residues      1-475 ##label DED
      ##cross-references EMBL:Z49809; NID:g854459; PID:g854463; MIPS:YMR211w
      ##experimental_source strain AB972
GENETICS
   #map_position 13R
SUMMARY          #length 475  #molecular-weight 55312  #checksum 2719

  Query Match           65.3%;  Score 49;  DB 2;  Length 475;
  Best Local Similarity  60.0%;  Pred. No. 2.52e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    321 YLTTAITLGY 330
          ||:|| :| |
Qy      3 YLSTASSLEY 12


RESULT   14
ENTRY            ACGAE      #type complete
TITLE            glutamate receptor precursor - great pond snail
ORGANISM         #formal_name Lymnaea stagnalis #common_name great pond snail
DATE             31-Mar-1992 #sequence_revision 28-Oct-1994 #text_change
                   05-Sep-1997
ACCESSIONS       S18443; S15681
REFERENCE        S18443
   #authors      Hutton, M.L.; Harvey, R.J.; Barnard, E.A.; Darlison, M.G.
   #journal      FEBS Lett. (1991) 292:111-114
   #title        Cloning of a cDNA that encodes an invertebrate glutamate
                   receptor subunit.
   #cross-references MUID:92070466
   #accession    S18443
      ##status       nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-917 ##label HUT
      ##cross-references EMBL:X60086
REFERENCE        S15681
   #authors      Hutton, M.L.; Bhandal, N.S.; Harvey, R.J.; Usherwood, P.N.R.;
                   Darlison, M.G.
   #submission   submitted to the EMBL Data Library, June 1991
   #description  PCR-mediated cloning of a cDNA that encodes a functional
                   molluscan glutamate receptor subunit.
   #accession    S15681
      ##molecule_type mRNA
      ##residues      1-362,'K',364-776,'S',778-845,'R',847-886,'S',888-917
                       ##label HU2
      ##cross-references EMBL:X60086; NID:g9628; PID:g9629
CLASSIFICATION   #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS         glycoprotein; ion channel; neurotransmitter receptor;
                   transmembrane protein

FEATURE
   1-19              #domain signal sequence #status predicted #label SIG\
   20-917            #product glutamate receptor #status predicted #label
                       MAT\
   429-853           #domain glutamate receptor homology #label GRH\
   559-578           #domain transmembrane #status predicted #label TM1\
   599-617           #domain transmembrane #status predicted #label TM2\
   628-646           #domain transmembrane #status predicted #label TM3\
   819-839           #domain transmembrane #status predicted #label TM4\
   62,95,121,125,229,
   251,261,272,418,
   419,424,491,881   #binding_site carbohydrate (Asn) (covalent) #status
                       predicted
SUMMARY          #length 917  #molecular-weight 103106  #checksum 8952

  Query Match           65.3%;  Score 49;  DB 1;  Length 917;
  Best Local Similarity  36.4%;  Pred. No. 2.52e+01;
  Matches     4;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db    729 AYLTESSTIDY 739
          :||: :|:::|
Qy      2 SYLSTASSLEY 12


RESULT   15
ENTRY            T03030     #type fragment
TITLE            hypothetical protein KIAA0365 - human (fragment)
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             23-Mar-1999 #sequence_revision 23-Mar-1999 #text_change
                   23-Mar-1999
ACCESSIONS       T03030
REFERENCE        Z14651
   #authors      Lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson,
                   A.W.; Burkhart-Schultz, K.; Gordon, L.; Kyle, A.; Ramirez,
                   M.; Stilwagen, S.; Phan, H.; Velasco, N.; Garnes, J.;
                   Danganan, L.; Poundstone, P.; Christensen, M.; Georgescu,
                   A.; Avila, J.; Liu, S.; Attix, C.; Andreise, T.; Trankheim,
                   M.; Amico-Keller, G.; Coefield, J.; Duarte, S.; Lucas, S.;
                   Bruce, R.; Thomas, P.; Quan, G.; Kronmiller, B.; Arellano,
                   A.; Montgomery, M.; Ow, D.; Nolan, M.
   #submission   submitted to the EMBL Data Library, March 1998
   #description  Sequence analysis of an ~1 Mb region containing the MEF2B
                   gene in 19p12.
   #accession    T03030
      ##status       preliminary
      ##residues      1-949 ##label LAM
      ##cross-references EMBL:AC004447; NID:g2978446; PID:g2978447
SUMMARY          #length 949  #checksum 5364

  Query Match           65.3%;  Score 49;  DB 3;  Length 949;
  Best Local Similarity  60.0%;  Pred. No. 2.52e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    475 FLSDENSLEY 484
          :||  :||||
Qy      3 YLSTASSLEY 12


Search completed: Thu Sep  2 12:50:54 1999
Job time : 14 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:52:29 1999; MasPar time 1.40 Seconds
86.863 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-34
Description: (1-12) from US08599226.pep
Perfect Score: 75
Sequence: 1 ASYLSTASSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 16.154; Variance 53.667; scale 0.301

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	50	66.7	37	3	PCT-US96-0	Sequence 5, Applicatio	7.21e+01
2	50	66.7	117	3	PCT-US96-0	Sequence 11, Applicati	7.21e+01
3	50	66.7	158	2	US-08-828-	Sequence 3, Applicatio	7.21e+01
4	50	66.7	159	2	US-08-828-	Sequence 1, Applicatio	7.21e+01
5	49	65.3	426	3	PCT-US95-1	Sequence 2, Applicatio	9.07e+01
6	49	65.3	426	1	US-08-336-	Sequence 2, Applicatio	9.07e+01
7	48	64.0	23	2	US-08-985-	Sequence 11, Applicati	1.14e+02
8	48	64.0	290	3	PCT-US95-1	Sequence 27, Applicati	1.14e+02
9	48	64.0	290	2	US-08-420-	Sequence 27, Applicati	1.14e+02
10	48	64.0	445	2	US-08-985-	Sequence 2, Applicatio	1.14e+02
11	48	64.0	1026	3	PCT-US95-0	Sequence 95, Applicati	1.14e+02
12	48	64.0	1026	2	US-08-268-	Sequence 95, Applicati	1.14e+02
13	48	64.0	1026	1	US-07-998-	Sequence 95, Applicati	1.14e+02
14	48	64.0	1026	1	US-08-453-	Sequence 95, Applicati	1.14e+02
15	48	64.0	1026	1	US-08-453-	Sequence 95, Applicati	1.14e+02
16	48	64.0	1026	3	PCT-US93-1	Sequence 95, Applicati	1.14e+02
17	48	64.0	1203	3	PCT-US95-0	Sequence 103, Applicat	1.14e+02
18	48	64.0	1203	3	PCT-US93-1	Sequence 103, Applicat	1.14e+02
19	48	64.0	1203	1	US-07-998-	Sequence 103, Applicat	1.14e+02
20	48	64.0	1203	2	US-08-268-	Sequence 103, Applicat	1.14e+02
21	48	64.0	1203	1	US-08-453-	Sequence 103, Applicat	1.14e+02
22	48	64.0	1203	1	US-08-453-	Sequence 103, Applicat	1.14e+02
23	48	64.0	1255	2	US-08-468-	Sequence 68, Applicati	1.14e+02
24	48	64.0	1255	2	US-08-625-	Sequence 2, Applicatio	1.14e+02
25	48	64.0	1255	2	US-08-484-	Sequence 8, Applicatio	1.14e+02
26	48	64.0	1255	2	US-08-486-	Sequence 68, Applicati	1.14e+02
27	48	64.0	1255	2	US-08-356-	Sequence 2, Applicatio	1.14e+02
28	48	64.0	1255	2	US-08-414-	Sequence 68, Applicati	1.14e+02
29	48	64.0	1255	1	US-08-467-	Sequence 68, Applicati	1.14e+02
30	47	62.7	20	2	US-08-468-	Sequence 2, Applicatio	1.43e+02
31	47	62.7	21	2	US-08-787-	Sequence 33, Applicati	1.43e+02
32	47	62.7	170	2	US-08-327-	Sequence 1, Applicatio	1.43e+02
33	47	62.7	509	3	PCT-US95-0	Sequence 8, Applicatio	1.43e+02
34	46	61.3	168	4	5194425-4	Patent No. 5194425.	1.79e+02
35	46	61.3	209	2	US-08-771-	Sequence 2, Applicatio	1.79e+02
36	46	61.3	345	2	US-08-446-	Sequence 40, Applicati	1.79e+02
37	46	61.3	384	1	US-08-457-	Sequence 5, Applicatio	1.79e+02
38	46	61.3	509	3	PCT-US95-0	Sequence 4, Applicatio	1.79e+02
39	46	61.3	824	2	US-08-785-	Sequence 7, Applicatio	1.79e+02
40	46	61.3	824	2	US-08-816-	Sequence 52, Applicati	1.79e+02
41	45	60.0	23	2	US-08-985-	Sequence 17, Applicati	2.23e+02
42	45	60.0	460	2	US-08-714-	Sequence 7, Applicatio	2.23e+02
43	45	60.0	505	1	US-08-631-	Sequence 2, Applicatio	2.23e+02
44	45	60.0	505	2	US-08-936-	Sequence 2, Applicatio	2.23e+02
45	45	60.0	505	2	US-08-829-	Sequence 2, Applicatio	2.23e+02

ALIGNMENTS

```
RESULT    1
ID   PCT-US96-08730-5     STANDARD;     PRT;   37 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application PC/TUS9608730
XX
CC    Sequence 5, Application PC/TUS9608730
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cassels, Frederick
CC       APPLICANT:  Anderson, Jeffrey
CC       APPLICANT:  Carter, John Mark
CC       TITLE OF INVENTION:  Methods of Raising Antibodies Against E.
CC       TITLE OF INVENTION:  Coli of the Family CSF-CFA./1
CC       NUMBER OF SEQUENCES:  15
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Glenna Hendricks
CC         STREET:  P.O. Box 2509
CC         CITY:  Fairfax
CC         STATE:  VA
CC         COUNTRY:  USA
CC         ZIP:  22031
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US96/08730
CC         FILING DATE:  03-JUN-1996
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Hendricks, Glenna
CC         REGISTRATION NUMBER:  32,535
CC         REFERENCE/DOCKET NUMBER:  PCT/US96/08730
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703) 591-4470
CC         TELEFAX:  (703) 591-4428
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  37 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
```

```
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  internal
SQ   SEQUENCE   37 AA; 3864 MW; 7776 CN;

  Query Match           66.7%;  Score 50;  DB 3;  Length 37;
  Best Local Similarity  58.3%;  Pred. No. 7.21e+01;
  Matches     7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db     22 GSYLPTAVELTY 33
          :|||:||  | |
Qy      1 ASYLSTASSLEY 12


RESULT    2
ID   PCT-US96-08730-11     STANDARD;     PRT;   117 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 11, Application PC/TUS9608730
XX
CC    Sequence 11, Application PC/TUS9608730
CC     GENERAL INFORMATION:
CC       APPLICANT:  Cassels, Frederick
CC       APPLICANT:  Anderson, Jeffrey
CC       APPLICANT:  Carter, John Mark
CC       TITLE OF INVENTION:  Methods of Raising Antibodies Against E.
CC       TITLE OF INVENTION:  Coli of the Family CSF-CFA./1
CC       NUMBER OF SEQUENCES:  15
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Glenna Hendricks
CC         STREET:  P.O. Box 2509
CC         CITY:  Fairfax
CC         STATE:  VA
CC         COUNTRY:  USA
CC         ZIP:  22031
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US96/08730
CC         FILING DATE:  03-JUN-1996
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Hendricks, Glenna
CC         REGISTRATION NUMBER:  32,535
CC         REFERENCE/DOCKET NUMBER:  PCT/US96/08730
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (703) 591-4470
CC         TELEFAX:  (703) 591-4428
CC     INFORMATION FOR SEQ ID NO:  11:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  117 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  internal
SQ   SEQUENCE   117 AA; 12389 MW; 76297 CN;

  Query Match           66.7%;  Score 50;  DB 3;  Length 117;
  Best Local Similarity  58.3%;  Pred. No. 7.21e+01;
  Matches     7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db     22 GSYLPTAVELTY 33
          :|||:||  | |
Qy      1 ASYLSTASSLEY 12


RESULT    3
ID   US-08-828-832-3     STANDARD;     PRT;   158 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08828832
XX
CC    Sequence 3, Application US/08828832
CC    Patent No. 5827711
CC     GENERAL INFORMATION:
CC       APPLICANT:  Lal, Preeti
CC       APPLICANT:  Shah, Purvi
CC       TITLE OF INVENTION:  NOVEL SUCCINATE DEHYDROGENASE SUBUNIT
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/828,832
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0250 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  158 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  1575011
SQ   SEQUENCE   158 AA; 17096 MW; 135076 CN;

  Query Match           66.7%;  Score 50;  DB 2;  Length 158;
  Best Local Similarity  41.7%;  Pred. No. 7.21e+01;
  Matches     5;  Conservative    5;  Mismatches   2;  Indels   0;  Gaps   0;

Db     81 AAYLNPCSAMDY 92
          |:||:  |:::|
Qy      1 ASYLSTASSLEY 12


RESULT    4
ID   US-08-828-832-1     STANDARD;     PRT;   159 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08828832
```

```
XX
CC    Sequence 1, Application US/08828832
CC    Patent No. 5827711
CC     GENERAL INFORMATION:
CC       APPLICANT:  Lal, Preeti
CC       APPLICANT:  Shah, Purvi
CC       TITLE OF INVENTION:  NOVEL SUCCINATE DEHYDROGENASE SUBUNIT
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/828,832
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0250 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  159 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  Consensus
CC         CLONE:  2454416
SQ   SEQUENCE   159 AA; 17043 MW; 129477 CN;

  Query Match           66.7%;  Score 50;  DB 2;  Length 159;
  Best Local Similarity  41.7%;  Pred. No. 7.21e+01;
  Matches     5;  Conservative    5;  Mismatches   2;  Indels   0;  Gaps   0;

Db     82 AAYLNPCSAMDY 93
          |:||:  |:::|
Qy      1 ASYLSTASSLEY 12


RESULT    5
ID   PCT-US95-13795-2     STANDARD;     PRT;   426 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9513795
XX
CC     Sequence 2, Application PC/TUS9513795
CC      GENERAL INFORMATION:
CC       APPLICANT:  HOLLIS, GREGORY F.
CC       APPLICANT:  PATEL, MAYUR D.
CC       TITLE OF INVENTION:  DNA ENCODING CANINE IMMUNOGLOBULINS
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  126 E. LINCOLN AVENUE; P.O. BOX 2000
CC         CITY:  RAHWAY
CC         STATE:  NEW JERSEY
CC         COUNTRY:  USA
CC         ZIP:  07065-0907
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/13795
CC         FILING DATE:
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
CC         REFERENCE/DOCKET NUMBER:  19211Y
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (908) 594-6734
CC         TELEFAX:  (908) 594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  426 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   426 AA; 47234 MW; 1032622 CN;

  Query Match           65.3%;  Score 49;  DB 3;  Length 426;
  Best Local Similarity  54.5%;  Pred. No. 9.07e+01;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    214 TSYLSPPSPLD 224
          :|||| :|:|:
Qy      1 ASYLSTASSLE 11


RESULT    6
ID   US-08-336-583-2     STANDARD;     PRT;   426 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08336583
XX
CC     Sequence 2, Application US/08336583
CC     Patent No. 5629415
CC      GENERAL INFORMATION:
CC       APPLICANT:  HOLLIS, GREGORY F.
CC       APPLICANT:  PATEL, MAYUR D.
CC       TITLE OF INVENTION:  DNA ENCODING CANINE IMMUNOGLOBULIN E
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  CHRISTINE E. CARTY
CC         STREET:  126 E. LINCOLN AVENUE
CC         CITY:  RAHWAY
CC         STATE:  NEW JERSEY
CC         COUNTRY:  USA
CC         ZIP:  07065-0900
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/336,583
CC         FILING DATE:  09-NOV-1994
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  CARTY, CHRISTINE E.
CC         REGISTRATION NUMBER:  36,099
```

```
CC         REFERENCE/DOCKET NUMBER:  19211
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (908) 594-6734
CC         TELEFAX:  (908) 594-4720
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  426 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   426 AA; 47234 MW; 1032622 CN;

  Query Match           65.3%;  Score 49;  DB 1;  Length 426;
  Best Local Similarity  54.5%;  Pred. No. 9.07e+01;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    214 TSYLSPPSPLD 224
          :|||| :|:|:
Qy      1 ASYLSTASSLE 11


RESULT    7
ID   US-08-985-090-11     STANDARD;     PRT;   23 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 11, Application US/08985090
XX
CC    Sequence 11, Application US/08985090
CC    Patent No. 5885893
CC     GENERAL INFORMATION:
CC       APPLICANT:  Andrew D.J. Goodearl
CC       TITLE OF INVENTION:  MUSCARINIC RECEPTORS AND USES THEREFOR
CC       NUMBER OF SEQUENCES:  28
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  LAHIVE & COCKFIELD, LLP
CC         STREET:  28 State Street
CC         CITY:  Boston
CC         STATE:  Massachusetts
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/985,090
CC         FILING DATE:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Jean M. Silveri
CC         REGISTRATION NUMBER:  39,030
CC         REFERENCE/DOCKET NUMBER:  MNI-032
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617)227-7400
CC         TELEFAX:  (617)742-4214
CC     INFORMATION FOR SEQ ID NO:  11:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  23 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   23 AA; 2719 MW; 3615 CN;

  Query Match           64.0%;  Score 48;  DB 2;  Length 23;
  Best Local Similarity  60.0%;  Pred. No. 1.14e+02;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db      2 FLITASTLEF 11
          :| |||:||:
Qy      3 YLSTASSLEY 12


RESULT    8
ID   PCT-US95-10194-27     STANDARD;     PRT;   290 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 27, Application PC/TUS9510194
XX
CC    Sequence 27, Application PC/TUS9510194
CC     GENERAL INFORMATION:
CC       APPLICANT:  The Trustees of Columbia University in the City of New York
CC       APPLICANT:  City
CC       TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
CC       TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
CC       NUMBER OF SEQUENCES:  45
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham LLP
CC         STREET:  1185 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/10194
CC         FILING DATE:
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White, John P.
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  45185-C-PCT/JPW/MSC
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 278-0400
CC         TELEFAX:  (212) 391-0525
CC     INFORMATION FOR SEQ ID NO:  27:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  290 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   290 AA; 32186 MW; 443934 CN;

  Query Match           64.0%;  Score 48;  DB 3;  Length 290;
  Best Local Similarity  54.5%;  Pred. No. 1.14e+02;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    100 PYLVTPSSIEF 110
          :|| |:||:|:
Qy      2 SYLSTASSLEY 12


RESULT    9
ID   US-08-420-235B-27     STANDARD;     PRT;   290 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 27, Application US/08420235B
XX
CC    Sequence 27, Application US/08420235B
CC    Patent No. 5801042
```

```
CC     GENERAL INFORMATION:
CC       APPLICANT:  Chang, Yuan
CC       APPLICANT:  Moore, Patrick S.
CC       TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
CC       TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
CC       NUMBER OF SEQUENCES:  47
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham LLP
CC         STREET:  1185 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/420,235B
CC         FILING DATE:
CC         CLASSIFICATION:  424
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White, John P.
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  45185-B
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 278-0400
CC         TELEFAX:  (212) 391-0525
CC     INFORMATION FOR SEQ ID NO:  27:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  290 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   290 AA; 32186 MW; 443934 CN;

  Query Match           64.0%;  Score 48;  DB 2;  Length 290;
  Best Local Similarity  54.5%;  Pred. No. 1.14e+02;
  Matches     6;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    100 PYLVTPSSIEF 110
          :|| |:||:|:
Qy      2 SYLSTASSLEY 12


RESULT   10
    US-08-985-090-2     STANDARD;     PRT;   445 AA.

AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08985090
XX
CC    Sequence 2, Application US/08985090
CC    Patent No. 5885893
CC     GENERAL INFORMATION:
CC       APPLICANT:  Andrew D.J. Goodearl
CC       TITLE OF INVENTION:  MUSCARINIC RECEPTORS AND USES THEREFOR
CC       NUMBER OF SEQUENCES:  28
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  LAHIVE & COCKFIELD, LLP
CC         STREET:  28 State Street
CC         CITY:  Boston
CC         STATE:  Massachusetts
CC         COUNTRY:  USA
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/985,090
CC         FILING DATE:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Jean M. Silveri
CC         REGISTRATION NUMBER:  39,030
CC         REFERENCE/DOCKET NUMBER:  MNI-032
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617)227-7400
CC         TELEFAX:  (617)742-4214
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  445 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   445 AA; 48671 MW; 1062278 CN;

  Query Match           64.0%;  Score 48;  DB 2;  Length 445;
  Best Local Similarity  60.0%;  Pred. No. 1.14e+02;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    198 FLITASTLEF 207
          :| |||:||:
Qy      3 YLSTASSLEY 12


RESULT   11
ID   PCT-US95-08071-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application PC/TUS9508071
XX
CC    Sequence 95, Application PC/TUS9508071
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  6300 Sears Tower, 233 S. Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/08071
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/12588
CC         FILING DATE:  23 DEC 1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/998,003
CC         FILING DATE:  29 DEC 1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Noland, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32149
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
```

```
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           64.0%;  Score 48;  DB 3;  Length 1026;
  Best Local Similarity  40.0%;  Pred. No. 1.14e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          :| |:::|:|
Qy      3 YLSTASSLEY 12


RESULT   12
    US-08-268-161A-95     STANDARD;     PRT;   1026 AA.

AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application US/08268161A
XX
CC    Sequence 95, Application US/08268161A
CC    Patent No. 5798224
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  233 South Wacker, 6300 Sears Tower
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/268,161A
CC         FILING DATE:  June 27, 1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Young J. Suh
CC         REGISTRATION NUMBER:  P-41,337
CC         REFERENCE/DOCKET NUMBER:  27866/32149
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           64.0%;  Score 48;  DB 2;  Length 1026;
  Best Local Similarity  40.0%;  Pred. No. 1.14e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          :| |:::|:|
Qy      3 YLSTASSLEY 12


RESULT   13
ID   US-07-998-003A-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application US/07998003A
XX
CC    Sequence 95, Application US/07998003A
CC    Patent No. 5643781
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Bicknell
CC         STREET:  20 South Clark Street
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60603
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/998,003A
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5643781and, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  30903
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/346-5750
CC         TELEFAX:  312/984-9740
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           64.0%;  Score 48;  DB 1;  Length 1026;
  Best Local Similarity  40.0%;  Pred. No. 1.14e+02;
  Matches    4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          :| |:::|:|
Qy      3 YLSTASSLEY 12


RESULT   14
ID   US-08-453-695A-95     STANDARD;     PRT;   1026 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 95, Application US/08453695A
XX
CC    Sequence 95, Application US/08453695A
CC    Patent No. 5708143
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
```

```
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  115
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
CC         ADDRESSEE:  Borun
CC         STREET:  233 South Wacker, 6300 Sears Tower
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60606
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/453,695A
CC         FILING DATE:
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5708143and, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32658
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           64.0%;  Score 48;  DB 1;  Length 1026;
  Best Local Similarity  40.0%;  Pred. No. 1.14e+02;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          :| |:::|:|
Qy      3 YLSTASSLEY 12


RESULT   15
ID   US-08-453-274B-95     STANDARD;     PRT;   1026 AA.
XX
AC   XXXXXX
XX
DT
XX
DE   Sequence 95, Application US/08453274B
XX
CC    Sequence 95, Application US/08453274B
CC    Patent No. 5663300
CC     GENERAL INFORMATION:
CC       APPLICANT:  Suzuki, Shintaro
CC       TITLE OF INVENTION:  Protocadherin Materials and Methods
CC       NUMBER OF SEQUENCES:  107
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
CC         STREET:  6300 Sears Tower, 233 South Wacker Drive
CC         CITY:  Chicago
CC         STATE:  Illinois
CC         COUNTRY:  United States of America
CC         ZIP:  60606-6402
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/453,274B
CC         FILING DATE:  30-MAY-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  No. 5663300and, Greta E.
CC         REGISTRATION NUMBER:  35,302
CC         REFERENCE/DOCKET NUMBER:  32660
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  312/474-6300
CC         TELEFAX:  312/474-0448
CC         TELEX:  25-3856
CC     INFORMATION FOR SEQ ID NO:  95:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1026 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1026 AA; 111270 MW; 5611711 CN;

  Query Match           64.0%;  Score 48;  DB 1;  Length 1026;
  Best Local Similarity  40.0%;  Pred. No. 1.14e+02;
  Matches     4;  Conservative    5;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 FLQTTTPLDY 432
          :| |:::|:|
Qy      3 YLSTASSLEY 12


Search completed: Thu Sep  2 12:52:37 1999
Job time : 8 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 12:50:01 1999; MasPar time 3.46 Seconds
73.774 Million cell updates/sec
Tabular output not generated.

Title: >US-08-599-226-34
Description: (1-12) from US08599226.pep
Perfect Score: 75
Sequence: 1 ASYLSTASSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.323; Variance 56.648; scale 0.306

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	75	100.0	12	27	W27593	Anti-TNF-alpha antibo	4.20e-01
2	72	96.0	12	27	W27588	Anti-TNF-alpha antibo	8.89e-01
3	70	93.3	121	27	W27569	Anti-TNF-alpha antibo	1.46e+00
4	68	90.7	12	27	W27592	Anti-TNF-alpha antibo	2.39e+00
5	67	89.3	12	27	W27591	Anti-TNF-alpha antibo	3.06e+00
6	64	85.3	12	27	W27590	Anti-TNF-alpha antibo	6.36e+00
7	60	80.0	12	27	W27586	Anti-TNF-alpha antibo	1.66e+01
8	60	80.0	12	27	W27589	Anti-TNF-alpha antibo	1.66e+01
9	60	80.0	12	27	W27587	Anti-TNF-alpha antibo	1.66e+01
10	58	77.3	12	27	W27563	Anti-TNF-alpha antibo	2.68e+01
11	58	77.3	12	27	W27594	Anti-TNF-alpha antibo	2.68e+01
12	57	76.0	159	29	W55260	H. pylori ORF 06cp306	3.39e+01
13	57	76.0	177	29	W55472	H. pylori ORF 06cp306	3.39e+01
14	57	76.0	526	29	W55688	H. pylori ORF 09cp107	3.39e+01
15	53	70.7	875	28	W34987	Bankia gouldi endoglu	8.60e+01
16	50	66.7	37	22	W17907	Peptide CS4 from dena	1.71e+02
17	50	66.7	37	26	W24223	Peptide fragment from	1.71e+02
18	50	66.7	117	22	W17913	Peptide CS4 from dena	1.71e+02
19	50	66.7	159	37	W80222	Human succinate-ubiqu	1.71e+02
20	50	66.7	2799	39	W81867	Human tumour suppress	1.71e+02
21	49	65.3	417	26	W23067	Canine IgE heavy chai	2.14e+02
22	48	64.0	31	20	W03903	Glucagon like peptide	2.68e+02
23	48	64.0	166	32	W59345	Human ErbB2 domain 1	2.68e+02
24	48	64.0	191	32	W59354	Human ErbB2 protein f	2.68e+02
25	48	64.0	290	18	R97837	Kaposi's sarcoma asso	2.68e+02
26	48	64.0	290	17	R93613	Kaposi's sarcoma asso	2.68e+02
27	48	64.0	409	34	W71495	Helicobacter polypept	2.68e+02
28	48	64.0	771	39	W89589	Aspergillus oryzae di	2.68e+02
29	48	64.0	782	23	W19764	Her2-GM-CSF immunosti	2.68e+02
30	48	64.0	1026	17	R87146	Protocadherin clone 4	2.68e+02
31	48	64.0	1026	11	R58906	Human protocadherin-4	2.68e+02
32	48	64.0	1203	17	R87152	Alternatively spliced	2.68e+02
33	48	64.0	1203	11	R58911	Product of alternativ	2.68e+02
34	48	64.0	1255	19	W01111	HER-2/neu protein.	2.68e+02
35	48	64.0	1433	8	R39568	Sequence of c-erbB-2	2.68e+02
36	47	62.7	20	39	W73609	Human myelin basic pr	3.34e+02
37	47	62.7	20	19	R95334	MBP-1 (11-30).	3.34e+02
38	47	62.7	21	27	W37552	Human myelin basic pr	3.34e+02
39	47	62.7	21	19	R95336	MBP-1.2 (11-31).	3.34e+02
40	47	62.7	167	9	R48595	Myelin basic protein.	3.34e+02
41	47	62.7	170	9	R48592	Human myelin basic pr	3.34e+02
42	47	62.7	171	18	R99580	Human myelin basic pr	3.34e+02
43	47	62.7	375	19	W06104	PM4 chimera (delta PL	3.34e+02
44	47	62.7	385	19	W06102	MP3 chimera (MBP21.5-	3.34e+02
45	47	62.7	1464	38	W79294	An antigen from dermo	3.34e+02

ALIGNMENTS

RESULT 1
ID W27593 standard; peptide; 12 AA.
AC W27593;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 75; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          100.0%;  Score 75;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 4.20e-01;
  Matches    12;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstassley 12
          ||||||||||||
Qy      1 ASYLSTASSLEY 12


RESULT    2
ID   W27588 standard; peptide; 12 AA.
AC   W27588;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          96.0%;  Score 72;  DB 27;  Length 12;
  Best Local Similarity 83.3%;  Pred. No. 8.89e-01;
  Matches    10;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssldy 12
          ||||||:|||:|
Qy      1 ASYLSTASSLEY 12


RESULT    3
ID   W27569 standard; Protein; 121 AA.
AC   W27569;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain variable region.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody;
KW   heavy chain; variable region; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
DR   N-PSDB; T88404.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 16; Page 76; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain variable region.
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   121 AA;

  Query Match          93.3%;  Score 70;  DB 27;  Length 121;
  Best Local Similarity 90.9%;  Pred. No. 1.46e+00;
  Matches    10;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    100 sylstassldy 110
          |||||||||:|
Qy      2 SYLSTASSLEY 12


RESULT    4
ID   W27592 standard; peptide; 12 AA.
AC   W27592;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
```

KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 90.7%; Score 68; DB 27; Length 12;
Best Local Similarity 83.3%; Pred. No. 2.39e+00;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 asflstsssley 12
||:|||:|||||
Qy 1 ASYLSTASSLEY 12

RESULT 5
ID W27591 standard; peptide; 12 AA.
AC W27591;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 89.3%; Score 67; DB 27; Length 12;
Best Local Similarity 83.3%; Pred. No. 3.06e+00;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 asylstssslhy 12
||||||:||| |
Qy 1 ASYLSTASSLEY 12

RESULT 6
ID W27590 standard; peptide; 12 AA.
AC W27590;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).

CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 85.3%; Score 64; DB 27; Length 12;
Best Local Similarity 75.0%; Pred. No. 6.36e+00;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1 asylstsfsldy 12
|||||| :||:|
Qy 1 ASYLSTASSLEY 12

RESULT 7
ID W27586 standard; peptide; 12 AA.
AC W27586;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 72; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 80.0%; Score 60; DB 27; Length 12;
Best Local Similarity 81.8%; Pred. No. 1.66e+01;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 asylstsssld 11
|||||| :|||:
Qy 1 ASYLSTASSLE 11

RESULT 8
ID W27589 standard; peptide; 12 AA.
AC W27589;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 80.0%; Score 60; DB 27; Length 12;
Best Local Similarity 81.8%; Pred. No. 1.66e+01;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 asylstsssld 11
|||||| :|||:
Qy 1 ASYLSTASSLE 11

```
RESULT    9
ID   W27587 standard; peptide; 12 AA.
AC   W27587;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           80.0%;  Score 60;  DB 27;  Length 12;
  Best Local Similarity  81.8%;  Pred. No. 1.66e+01;
  Matches     9;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssld 11
          ||||||:|||:
Qy      1 ASYLSTASSLE 11


RESULT   10
ID   W27563 standard; peptide; 12 AA.
AC   W27563;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 12
FT                   /label= Tyr, Asn
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 9; Page 65; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           77.3%;  Score 58;  DB 27;  Length 12;
  Best Local Similarity  90.0%;  Pred. No. 2.68e+01;
  Matches     9;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      2 sylstassld 11
          |||||||||:
Qy      2 SYLSTASSLE 11


RESULT   11
ID   W27594 standard; peptide; 12 AA.
AC   W27594;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
```

PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Disclosure; Page 75; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 77.3%; Score 58; DB 27; Length 12;
Best Local Similarity 90.0%; Pred. No. 2.68e+01;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 2 sylstassld 11
|||||||||:
Qy 2 SYLSTASSLE 11

RESULT 12
ID W55260 standard; Protein; 159 AA.
AC W55260;
DT 02-JUL-1998 (first entry)
DE H. pylori ORF 06cp30603orf16 protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR N-PSDB; V24669.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claim 14; Page 502; 1145pp; English.
CC This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence 159 AA;

Query Match 76.0%; Score 57; DB 29; Length 159;
Best Local Similarity 80.0%; Pred. No. 3.39e+01;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 60 ylslasaley 69
||| ||:|||
Qy 3 YLSTASSLEY 12

RESULT 13
ID W55472 standard; Protein; 177 AA.
AC W55472;
DT 24-JUN-1998 (first entry)
DE H. pylori ORF 06cp30603_23452_c3_80 inner membrane protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR N-PSDB; V24881.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 14,80; Page 679; 1145pp; English.
CC This sequence is a H. pylori cell envelope inner membrane protein
CC involved in cofactor metabolism.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence 177 AA;

Query Match 76.0%; Score 57; DB 29; Length 177;
Best Local Similarity 80.0%; Pred. No. 3.39e+01;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Db      60 ylslasaley 69
           ||| ||:|||
Qy       3 YLSTASSLEY 12
```

RESULT 14
ID W55688 standard; Protein; 526 AA.
AC W55688;
DT 07-JUL-1998 (first entry)
DE H. pylori ORF 09cp10713_23452_c3_195 inner membrane protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR N-PSDB; V25097.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 14,80; Pages 947-948; 1145pp; English.
CC This sequence is a H. pylori cell envelope inner membrane protein
CC involved in cofactor metabolism.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence 526 AA;

```
  Query Match           76.0%;  Score 57;  DB 29;  Length 526;
  Best Local Similarity 80.0%;  Pred. No. 3.39e+01;
  Matches     8;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db      60 ylslasaley 69
           ||| ||:|||
Qy       3 YLSTASSLEY 12
```

RESULT 15
ID W34987 standard; Protein; 875 AA.
AC W34987;
DT 21-MAY-1998 (first entry)
DE Bankia gouldi endoglucanase.
KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase.
OS Bankia gouldi mix (Clone 37GP2).
PN WO9744361-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08793.
PR 22-MAY-1996; US-651572.
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
PI Lam DE, Mathur EJ;
DR WPI; 98-018435/02.
DR N-PSDB; T94195.
PT Endoglucanase(s), preferably form archael bacterium, AEPII 1a -
PT useful to degrade carboxymethylcellulose and hydrolyse of
PT beta-1,4-glycosidic bonds in cellulose
PS Claim 1; Fig 1C; 164pp; English.
CC This protein comprises an endoglucanase of Bankia gouldi mix (clone
CC 37GP2) that is capable of degrading carboxymethylcellulose and of
CC hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has
CC homology to an endoglucanase of archaebacterium AEPIIla (see
CC W34985). It can be produced from native cells or from recombinant
CC host cells, especially prokaryotic host cells transformed with a
CC plasmid or virus-derived vector including the endoglucanase DNA
CC (see T94195). 24 Endoglucanases (see W34986-W35008) are claimed.
CC They can be used to degrade cellulose for the conversion of plant
CC biomass into fuels and chemicals, for use in detergents, textiles,
CC animal feed, waste treatment, and in the fruit juice and brewing
CC industries for the clarification and extraction of juices.
SQ Sequence 875 AA;

```
  Query Match           70.7%;  Score 53;  DB 28;  Length 875;
  Best Local Similarity 58.3%;  Pred. No. 8.60e+01;
  Matches     7;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db     618 psyltvdsslty 629
           :|||:  ||| |
Qy       1 ASYLSTASSLEY 12
```

Search completed: Thu Sep 2 12:50:23 1999
Job time : 22 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

n on: Thu Sep 2 12:46:15 1999; MasPar time 3.60 Seconds
70.804 Million cell updates/sec
bular output not generated.

Title: >US-08-599-226-33
Description: (1-12) from US08599226.pep
Perfect Score: 74
Sequence: 1 ASFLSTSSSLEY 12

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.466; Variance 53.940; scale 0.324

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	74	100.0	12	27	W27592	Anti-TNF-alpha antibo	3.21e-01
2	69	93.2	12	27	W27588	Anti-TNF-alpha antibo	1.19e+00
3	68	91.9	12	27	W27593	Anti-TNF-alpha antibo	1.54e+00
4	64	86.5	12	27	W27591	Anti-TNF-alpha antibo	4.31e+00
5	63	85.1	121	27	W27569	Anti-TNF-alpha antibo	5.56e+00
6	61	82.4	12	27	W27590	Anti-TNF-alpha antibo	9.22e+00
7	57	77.0	12	27	W27586	Anti-TNF-alpha antibo	2.50e+01
8	57	77.0	12	27	W27589	Anti-TNF-alpha antibo	2.50e+01
9	57	77.0	12	27	W27587	Anti-TNF-alpha antibo	2.50e+01
10	52	70.3	1026	17	R87146	Protocadherin clone 4	8.47e+01
11	52	70.3	1026	11	R58906	Human protocadherin-4	8.47e+01
12	52	70.3	1203	17	R87152	Alternatively spliced	8.47e+01
13	52	70.3	1203	11	R58911	Product of alternativ	8.47e+01
14	51	68.9	12	27	W27563	Anti-TNF-alpha antibo	1.08e+02
15	51	68.9	12	27	W27594	Anti-TNF-alpha antibo	1.08e+02
16	50	67.6	159	29	W55260	H. pylori ORF 06cp306	1.37e+02
17	50	67.6	177	29	W55472	H. pylori ORF 06cp306	1.37e+02
18	50	67.6	235	17	R91543	Pertussis toxin clone	1.37e+02
19	50	67.6	526	29	W55688	H. pylori ORF 09cp107	1.37e+02
20	49	66.2	114	12	R62919	Human cytomegalovirus	1.73e+02
21	49	66.2	509	15	R88465	Hamster scavenger rec	1.73e+02
22	48	64.9	159	37	W80222	Human succinate-ubiqu	2.19e+02
23	48	64.9	875	28	W34987	Bankia gouldi endoglu	2.19e+02
24	47	63.5	287	14	R73013	Functional equivalent	2.76e+02
25	46	62.2	218	21	R86811	Saccharomyces cerevis	3.48e+02
26	46	62.2	311	38	W89812	Protein encoded by cl	3.48e+02
27	46	62.2	411	21	R86810	Saccharomyces cerevis	3.48e+02
28	46	62.2	882	8	R43959	N-heparan sulphate su	3.48e+02
29	45	60.8	312	33	W65043	Forsythia pinoresinol	4.38e+02
30	45	60.8	312	33	W65038	Forsythia pinoresinol	4.38e+02
31	45	60.8	312	32	W65039	Forsythia pinoresinol	4.38e+02
32	45	60.8	312	32	W65042	Forsythia pinoresinol	4.38e+02
33	45	60.8	372	7	R37996	Sequence of human mac	4.38e+02
34	45	60.8	408	20	W10082	N-delta-3/C-delta-411	4.38e+02
35	45	60.8	450	38	W86335	Kidney injury associa	4.38e+02
36	45	60.8	519	20	W10072	N-delta 3 LCSF.	4.38e+02
37	45	60.8	519	20	W10084	N-delta-3/C-delta-221	4.38e+02
38	45	60.8	520	2	R12111	Human M-CSF derivativ	4.38e+02
39	45	60.8	536	20	W10066	Human long form CSF-1	4.38e+02
40	45	60.8	536	1	P80764	Deduced amino acid se	4.38e+02
41	45	60.8	536	24	W22613	Human colony stimulat	4.38e+02
42	45	60.8	554	1	P91898	Sequence of a macroph	4.38e+02
43	45	60.8	554	9	R48677	Human MCSF.	4.38e+02
44	45	60.8	554	2	P81040	Sequence of human col	4.38e+02
45	45	60.8	554	7	R37997	Sequence of human mac	4.38e+02

ALIGNMENTS

RESULT 1
ID W27592 standard; peptide; 12 AA.
AC W27592;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple

```
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match          100.0%;  Score 74;  DB 27;  Length 12;
  Best Local Similarity 100.0%;  Pred. No. 3.21e-01;
  Matches   12;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asflstsssley 12
          ||||||||||||
Qy      1 ASFLSTSSSLEY 12


RESULT    2
ID   W27588 standard; peptide; 12 AA.
AC   W27588;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           93.2%;  Score 69;  DB 27;  Length 12;
  Best Local Similarity  83.3%;  Pred. No. 1.19e+00;
  Matches   10;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstsssldy 12
          ||:|||||||:|
Qy      1 ASFLSTSSSLEY 12


RESULT    3
ID   W27593 standard; peptide; 12 AA.
AC   W27593;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 75; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           91.9%;  Score 68;  DB 27;  Length 12;
  Best Local Similarity  83.3%;  Pred. No. 1.54e+00;
  Matches   10;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 asylstassley 12
          ||:|||:|||||
Qy      1 ASFLSTSSSLEY 12


RESULT    4
ID   W27591 standard; peptide; 12 AA.
AC   W27591;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
```

KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 86.5%; Score 64; DB 27; Length 12;
Best Local Similarity 83.3%; Pred. No. 4.31e+00;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 asylstssslhy 12
||:|||||| |
Qy 1 ASFLSTSSSLEY 12

RESULT 5
ID W27569 standard; Protein; 121 AA.
AC W27569;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain variable region.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody;
KW heavy chain; variable region; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
DR N-PSDB; T88404.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 16; Page 76; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain variable region.
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 121 AA;

Query Match 85.1%; Score 63; DB 27; Length 121;
Best Local Similarity 72.7%; Pred. No. 5.56e+00;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 100 sylstassldy 110
|:|||:|||:|
Qy 2 SFLSTSSSLEY 12

RESULT 6
ID W27590 standard; peptide; 12 AA.
AC W27590;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 74; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).

CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 82.4%; Score 61; DB 27; Length 12;
Best Local Similarity 75.0%; Pred. No. 9.22e+00;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1 asylstsfsldy 12
||:|||| ||:|
Qy 1 ASFLSTSSSLEY 12

RESULT 7
ID W27586 standard; peptide; 12 AA.
AC W27586;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 72; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 77.0%; Score 57; DB 27; Length 12;
Best Local Similarity 81.8%; Pred. No. 2.50e+01;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 asylstsssld 11
||:|||||||:
Qy 1 ASFLSTSSSLE 11

RESULT 8
ID W27589 standard; peptide; 12 AA.
AC W27589;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Claim 20; Page 73; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 77.0%; Score 57; DB 27; Length 12;
Best Local Similarity 81.8%; Pred. No. 2.50e+01;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 asylstsssld 11
||:|||||||:
Qy 1 ASFLSTSSSLE 11

```
RESULT    9
ID   W27587 standard; peptide; 12 AA.
AC   W27587;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 20; Page 73; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
CC   pyrexia, periodontal disease, obesity and radiation toxicity. The
CC   Ab also inhibits TNF-alpha induced expression of endothelial cell
CC   leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC   endothelial cells (HUVEC).
SQ   Sequence   12 AA;

  Query Match           77.0%;  Score 57;  DB 27;  Length 12;
  Best Local Similarity  81.8%;  Pred. No. 2.50e+01;
  Matches    9;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 asylstsssld 11
           ||:|||||||:
Qy       1 ASFLSTSSSLE 11


RESULT   10
ID   R87146 standard; Protein; 1026 AA.
AC   R87146;
DT   29-AUG-1996  (first entry)
DE   Protocadherin clone 42.
KW   Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion; mouse;
KW   catenin; therapy; clone; frog; fruit fly.
OS   Homo sapiens.
PN   WO9600289-A1.
PD   04-JAN-1996.
PF   26-JUN-1995; U08071.
PR   27-JUN-1994; US-268161.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 96-068873/07.
DR   N-PSDB; T03621.
PT   Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT   pc5 - involved in cell-cell adhesion and regulation activities
PS   Example 3; Page 71-75; 146pp; English.
CC   This sequence represents a clone of the human protocadherin sequence,
CC   designated pc42.  The cDNA encoding this sequence was isolated after
CC   screening a human foetal brain cDNA library (contained within lambdaZapII
CC   vectors), with 32P labelled versions of the sequences represented by
CC   T03605 and T03606.  The cytoplasmic domain of cadherin interacts with the
CC   cytoskeleton through catenins and other cytoskeleton associated proteins.
CC   The cytoplasmic domain is not present in all cadherins, but in those
CC   which possess it, it is essential for the cadherins adhesive function.
CC   The cadherins which do not possess a cytoplasmic domain appear to
CC   function via a different method from those with a cytoplasmic domain.
CC   This protein sequence is involved in cell-cell adhesion.  This sequence
CC   may have regulatory functions in the cell, as well as the cell-cell
CC   adhesive properties.  Antibodies produced against this sequence are
CC   useful for modulating the binding activity of protocadherins, and can be
CC   used therapeutically.
SQ   Sequence   1026 AA;

  Query Match           70.3%;  Score 52;  DB 17;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 8.47e+01;
  Matches    5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db     423 flqtttpldy 432
           || |:::|:|
Qy       3 FLSTSSSLEY 12


RESULT   11
ID   R58906 standard; Protein; 1026 AA.
AC   R58906;
DT   17-APR-1995  (first entry)
DE   Human protocadherin-42.
KW   Cadherin; protocadherin; cell adhesion molecule.
OS   Homo sapiens.
PN   WO9414960-A.
PD   07-JUL-1994.
PF   23-DEC-1993; U12588.
PR   29-DEC-1992; US-998003.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 94-293849/36.
DR   N-PSDB; Q68997.
PT   Polynucleotide sequences encoding new proto:cadherins - useful
PT   for modulating natural binding and regulating activities.
PS   Claim 13; Page 66-70; 114pp; English.
CC   Two full length human cDNAs corresp. to the partial cDNAs of
CC   HUMAN-42 and HUMAN-43 (Q68981,Q68982) were isolated
CC   from human fetal brain cDNA library. Several overlapping cDNA
CC   clones were isolated with each probe including two cDNAs which
CC   contained the putative entire coding sequences of two novel
CC   proteins designated protocadherin-42 (pc42) and protocadherin-43
CC   (pc43). The DNA and deduced AA sequences of pc42 are in
CC   Q68997/R58906, while those of pc43 are in Q68998/R58907. The
CC   overall structures of pc42 and pc43 are similar to that of
CC   typical cadherins but they do have distinct features. Both lack
CC   the prosequences that are present in all known cadherin precursors.
CC   The extracellular domains of pc42 and pc43 are different in length
CC   and pc42 contains seven subdomains that closely resemble the
CC   typical cadherin extracellular subdomain while pc43 has six such
CC   domains. The sequences do not show any significant homology with
CC   those of known cadherins or cadherin-related proteins.
SQ   Sequence   1026 AA;

  Query Match           70.3%;  Score 52;  DB 11;  Length 1026;
  Best Local Similarity  50.0%;  Pred. No. 8.47e+01;
  Matches    5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db    423 flqtttpldy 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT   12
ID   R87152 standard; Protein; 1203 AA.
AC   R87152;
DT   29-AUG-1996  (first entry)
DE   Alternatively spliced pc42.
KW   Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion; mouse;
KW   catenin; therapy; clone; frog; fruit fly.
OS   Homo sapiens.
PN   WO9600289-A1.
PD   04-JAN-1996.
PF   26-JUN-1995; U08071.
PR   27-JUN-1994; US-268161.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 96-068873/07.
DR   N-PSDB; T03623.
PT   Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT   pc5 - involved in cell-cell adhesion and regulation activities
PS   Example 3; Page 90-96; 146pp; English.
CC   This sequence represents a possible alternatively spliced version of the
CC   clone of the human protocadherin sequence, designated pc42.  The cDNA
CC   encoding this sequence was isolated after screening a human foetal brain
CC   cDNA library (contained within lambdaZapII vectors), with 32P labelled
CC   versions of the sequences represented by T03605 and T03606.  The most
CC   abundant spliced version of pc42 is represented in T03621.  The
CC   cytoplasmic domain of cadherin interacts with the cytoskeleton through
CC   catenins and other cytoskeleton associated proteins.  The cytoplasmic
CC   domain is not present in all cadherins, but in those which possess it, it
CC   is essential for the cadherins adhesive function.  The cadherins which do
CC   not possess a cytoplasmic domain appear to function via a different
CC   method from those with a cytoplasmic domain.  This protein sequence is
CC   involved in cell-cell adhesion.  This sequence may have regulatory
CC   functions in the cell, as well as the cell-cell adhesive properties.
CC   Antibodies produced against this sequence are useful for modulating the
CC   binding activity of protocadherins, and can be used therapeutically.
SQ   Sequence   1203 AA;

  Query Match           70.3%;  Score 52;  DB 17;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 8.47e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 flqtttpldy 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT   13
ID   R58911 standard; Protein; 1203 AA.
AC   R58911;
DT   17-APR-1995  (first entry)
DE   Product of alternative splicing of human protocadherin-42 mRNA.
KW   Cadherin; protocadherin; cell adhesion molecule.
OS   Homo sapiens.
PN   WO9414960-A.
PD   07-JUL-1994.
PF   23-DEC-1993; U12588.
PR   29-DEC-1992; US-998003.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 94-293849/36.
DR   N-PSDB; Q68999.
PT   Polynucleotide sequences encoding new proto:cadherins - useful
PT   for modulating natural binding and regulating activities.
PS   Example; Page 84-89; 114pp; English.
CC   Two full length human cDNAs corresp. to the partial cDNAs of
CC   HUMAN-42 and HUMAN-43 (Q68981,Q68982) were isolated
CC   from human fetal brain cDNA library. Several overlapping cDNA
CC   clones were isolated with each probe including two cDNAs which
CC   contained the putative entire coding sequences of two novel
CC   proteins designated protocadherin-42 (pc42) and protocadherin-43
CC   (pc43). The DNA and deduced AA sequences of pc42 are in
CC   Q68997/R58906, while those of pc43 are in Q68998/R58907. Sequence
CC   analysis of various overlapping protocadherin cDNA clones revealed
CC   that some clones contd. unique sequences at the 3' end. The
CC   sequences forming the boundaries of the 3' end regions are
CC   consistent with the consensus sequence of mRNA splicing, suggesting
CC   that these clones may corresp. to alternatively spliced mRNAs.
CC   The DNA and AA sequences of one possible product of alternative
CC   splicing of pc42 mRNA are given in Q68999/R58911. The DNA and AA
CC   sequences of two possible products of alternative splicing of pc43
CC   mRNA are respectively presented in Q69000/R58912 and Q69001/R49144.
SQ   Sequence   1203 AA;

  Query Match           70.3%;  Score 52;  DB 11;  Length 1203;
  Best Local Similarity  50.0%;  Pred. No. 8.47e+01;
  Matches     5;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db    423 flqtttpldy 432
          || |:::|:|
Qy      3 FLSTSSSLEY 12


RESULT   14
ID   W27563 standard; peptide; 12 AA.
AC   W27563;
DT   19-MAR-1998  (first entry)
DE   Anti-TNF-alpha antibody heavy chain CDR3.
KW   Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW   heavy chain; complementarity determining region 3; inhibition;
KW   treatment; sepsis; disease; autoimmune disease; infectious disease;
KW   malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW   cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW   bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW   keloid formation; scar tissue formation; pyrexia; HUVEC;
KW   periodontal disease; obesity; radiation toxicity;
KW   endothelial cell leukocyte adhesion molecule-1;
KW   human umbilical vein endothelial cell.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 12
FT                   /label= Tyr, Asn
PN   WO9729131-A1.
PD   14-AUG-1997.
PF   10-FEB-1997; U02219.
PR   25-NOV-1996; US-031476.
PR   09-FEB-1996; US-599226.
PA   (BADI ) BASF AG.
PI   Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI   Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI   Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR   WPI; 97-415302/38.
PT   High affinity antibodies against human TNF alpha - useful to inhibit
PT   TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS   Claim 9; Page 65; 102pp; English.
CC   The present sequence is a novel anti-human tumour necrosis
CC   factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC   determining region 3 (CDR3).
CC   The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC   less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC   less (both determined by surface plasmon resonance), and
CC   neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC   L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC   inhibits TNF-alpha activity, can be used to treat sepsis,
CC   autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC   spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC   sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC   syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC   cardiac or inflammatory bone disorders, bone resorption disease,
CC   alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC   burns, reperfusion injury, keloid formation, scar tissue formation,
```

CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 68.9%; Score 51; DB 27; Length 12;
Best Local Similarity 70.0%; Pred. No. 1.08e+02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 2 sylstassld 11
|:|||:|||:
Qy 2 SFLSTSSSLE 11

RESULT 15
ID W27594 standard; peptide; 12 AA.
AC W27594;
DT 19-MAR-1998 (first entry)
DE Anti-TNF-alpha antibody heavy chain CDR3.
KW Human; tumour necrosis factor-alpha; TNF-alpha; antibody; CDR3;
KW heavy chain; complementarity determining region 3; inhibition;
KW treatment; sepsis; disease; autoimmune disease; infectious disease;
KW malignancy; pulmonary disorder; intestinal disorder; hepatitis;
KW cardiac disorder; inflammatory bone disorder; reperfusion injury;
KW bone resorption disease; coagulation disturbance; burn; ELAM-1;
KW keloid formation; scar tissue formation; pyrexia; HUVEC;
KW periodontal disease; obesity; radiation toxicity;
KW endothelial cell leukocyte adhesion molecule-1;
KW human umbilical vein endothelial cell.
OS Homo sapiens.
PN WO9729131-A1.
PD 14-AUG-1997.
PF 10-FEB-1997; U02219.
PR 25-NOV-1996; US-031476.
PR 09-FEB-1996; US-599226.
PA (BADI) BASF AG.
PI Allen DJ, Hoogenboom HRJM, Kaymakcalan Z, Labkovsky B,
PI Mankovich JA, McGuiness BT, Roberts AJ, Sakorafas P,
PI Salfeld JG, Schoenhaut D, Vaughan TJ, White M, Wilton AJ;
DR WPI; 97-415302/38.
PT High affinity antibodies against human TNF alpha - useful to inhibit
PT TNF alpha activity, e.g. to treat autoimmune diseases and cancer
PS Disclosure; Page 75; 102pp; English.
CC The present sequence is a novel anti-human tumour necrosis
CC factor-alpha (TNF-alpha) antibody (Ab) heavy chain complementarity
CC determining region 3 (CDR3).
CC The Ab dissociates from TNF-alpha with a Kd of 1x10 power -8 M or
CC less and has a Koff rate constant of 1x10 power -3 s power -1 or
CC less (both determined by surface plasmon resonance), and
CC neutralises human TNF-alpha cytotoxicity in a standard in vitro
CC L929 assay with an IC50 of 1x10 power -7 M or less. The Ab, which
CC inhibits TNF-alpha activity, can be used to treat sepsis,
CC autoimmune diseases, e.g. rheumatoid arthritis, rheumatoid
CC spondylitis, osteoarthritis, gouty arthritis, allergy, multiple
CC sclerosis, autoimmune diabetes, autoimmune uveitis or nephrotic
CC syndrome, infectious diseases, malignancy, pulmonary, intestinal,
CC cardiac or inflammatory bone disorders, bone resorption disease,
CC alcoholic, viral or fulminant hepatitis, coagulation disturbances,
CC burns, reperfusion injury, keloid formation, scar tissue formation,
CC pyrexia, periodontal disease, obesity and radiation toxicity. The
CC Ab also inhibits TNF-alpha induced expression of endothelial cell
CC leukocyte adhesion molecule-1 (ELAM-1) on human umbilical vein
CC endothelial cells (HUVEC).
SQ Sequence 12 AA;

Query Match 68.9%; Score 51; DB 27; Length 12;
Best Local Similarity 70.0%; Pred. No. 1.08e+02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 2 sylstassld 11
|:|||:|||:
Qy 2 SFLSTSSSLE 11

Search completed: Thu Sep 2 12:46:36 1999
Job time : 21 secs.